

10/070386

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 09:32:10 ; Search time 2707.4 Seconds  
(without alignments)  
15199.092 Million cell updates/sec

Title: US-10-070-386-1  
Perfect score: 1378  
Sequence: 1 gtcagcgggtggtgatat.....accatcaatcaactaaca 1378

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estm.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estro.\*
- 8: em\_hic.\*
- 9: gb\_est1.\*
- 10: gb\_est2.\*
- 11: gb\_hic.\*
- 12: gb\_est3.\*
- 13: gb\_est4.\*
- 14: gb\_est5.\*
- 15: em\_estfun.\*
- 16: em\_estom.\*
- 17: em\_gss\_hum.\*
- 18: em\_gss\_inv.\*
- 19: em\_gss\_pln.\*
- 20: em\_gss\_vrt.\*
- 21: em\_gss\_fun.\*
- 22: em\_gss\_mam.\*
- 23: em\_gss\_mus.\*
- 24: em\_gss\_pro.\*
- 25: em\_gss\_rod.\*
- 26: em\_gss\_phg.\*
- 27: em\_gss\_vrl.\*
- 28: gb\_gss1.\*
- 29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	43.8	3.2	885	13	BX425603
C 2	42.4	3.1	1201	13	BX376097
C 3	41.6	3.0	557	28	BH285801
C 4	39	2.8	1031	29	CNS05QOW

39	2.8	1201	13	BX381961
38.8	2.8	1205	28	CC364185
38.4	2.8	673	28	BZ246183
38.2	2.8	479	28	AQ018895
38.2	2.8	923	29	CNS04NFT
38	2.8	482	12	EG185531
38	2.8	1201	13	BX419736
37.8	2.7	1101	29	CNS000D1
37.8	2.7	1201	13	BX381961
37.6	2.7	688	9	AV384223
37.4	2.7	598	28	AZ960759
37.2	2.7	1110	13	BX447907
37.2	2.7	250	29	CE251889
37	2.7	590	29	CE571358
37	2.7	720	29	CE686685
37	2.7	801	12	BG534705
36.8	2.7	562	29	CE834892
36.8	2.7	1101	29	CNS016X9
36.6	2.7	1075	28	CC280186
36.4	2.6	1201	13	BX460908
36.2	2.6	457	12	B1741102
36.2	2.6	634	28	AZ037814
36	2.6	532	9	AL923574
36	2.6	551	28	AZ163313
36	2.6	621	12	BP504692
36	2.6	683	9	AA556936
36	2.6	1201	29	CNS0102D
35.8	2.6	432	12	B1502084
35.8	2.6	474	12	B1501745
35.8	2.6	592	28	AZ838544
35.8	2.6	610	28	AQ891964
35.8	2.6	845	28	BZ114424
35.8	2.6	1306	28	BZ565154
35.6	2.6	644	28	BH766160
35.6	2.6	748	12	B1886660
35.6	2.6	814	28	BH109031
35.6	2.6	959	29	CNS007MC
35.6	2.6	2077	11	AK080781
35.4	2.6	749	28	AQ895728
35.4	2.6	788	14	CF452751
35.4	2.6	885	13	BX425603

ALIGNMENTS

RESULT 1  
BX425603/c  
LOCUS  
DEFINITION BX425603 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone  
CLOBB022ZA07 3-PRIME, mRNA sequence.  
ACCESSION BX425603  
VERSION BX425603.1 GI:30770486  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Sukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 885)  
Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL Contact: Genoscope  
COMMENT Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ invitrogen Corporation 1600 Paraday Avenue Genoscope sequence ID : CLOBB022ZA07P1.  
FEATURES  
source location/Qualifiers  
1. .885  
/organism="Homo sapiens"

/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CL08B022A07"  
/tissue\_type="Homo sapiens NEUROBLASTOMA"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 3.2%; Score 43.8; DB 13; Length 885;  
Best Local Similarity 3.3%; Pred. No. 0.24;  
Matches 15; Conservative 146; Mismatches 293; Indels 0; Gaps 0;  
QY 92 AAGACTGAAGAGGACCAAGATATTCACATGCGATACGGTTCATCATGCTGTT 151  
DB MAMMM  
QY 152 CAAGAGAGACAGCTTCATCTACCTGCTGTTCCCTCTCTGTTGTACAGATACG 211  
DB MNN  
QY 212 GATGACACCCCGCAAGATTCCTGGAGTTCAAGAGGGTGTCTCTACGCGCATTT 271  
DB MAMMM  
QY 272 AGGTATAGTGCATPAGGCTTGACGTAAGCTGAAGCTGATTACGAGCATGAGACA 331  
DB MNN  
QY 332 GAAATAACAGGTTGTATGCTTCCCGTGTCTACTAAAGTGAATCAAGAGCAACA 391  
DB MAMMM  
QY 392 GCCGAAGAACCGATGCTGCTGAGGGTTCCTTAGAGTCTACATGTAACGTCAT 451  
DB MAMMM  
QY 452 GATAGAAACATCAATGCGCAATCAAGTTAGTATACCTGACGCTACATGCTTCTCCG 511  
DB MNN  
QY 512 GATCTGCTAAATATATGTCCTGCTCCGACT 545  
DB NANN

RESULT 2

LOCUS BX376097/c 1201 bp mRNA linear EST 08-MAY-2003  
DEFINITION BX376097 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
cDNA clone CSODC022TM12 5-PRIME, mRNA sequence.

ACCESSION BX376097  
VERSION BX376097.1 GI:30434756  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1201)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seq@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 2866.f  
Contact: Feng Liang Email: fliang@lifetech.com URL:  
http://fulllength.invitrogen.com/ invitrogen Corporation 1600

FEATURES

source

Paraday Avenue Genoscope sequence ID : CSODC022BG06QP1.  
Location/Qualifiers  
1. 1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODC022TM12"  
/tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 3.1%; Score 42.4; DB 13; Length 1201;  
Best Local Similarity 12.1%; Pred. No. 0.74; 385; Indels 1; Gaps 1;  
Matches 89; Conservative 260; Mismatches 385; Indels 1; Gaps 1;  
QY 297 GTAAGCTGAAGAGCTGATTACGAGCATGACACCAACGAAATATACACGGTTGTATGCGTTC 356  
DB GKRGGKGDAAADAAAAAAGKAAADAKGWAGKAAAGAAAGAAAKRKRAGKAG 1128  
QY 357 CGGTGCTTACTAAGATGATATCAAGAGCAACACAGCCGGAAGAACCGGATGCTGTCTGA 416  
DB KARGAAAGAGKAGAGAGGAGGARRKAKAGATARWGTAAAGKWRKARAAKKRKADAA 1068  
QY 417 GGGGTTCTTTAGAGTCTACATGCTAACGCTGATAGTAAACATCAATGCGCAATCA 476  
DB DGGKAAKAGKAGAGAGAGAGGAGGKGGKAAAKTWTMTMMMMNNNNNNNNNNNN 1008  
QY 477 AGTTAGTATACGCTACGCTACATCGCTTCTTCGGATCTTGCTCTAAAATATATGTCCT 536  
DB MAMMM 948  
QY 537 GTCGAACTCTGCTACTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 596  
DB MAMMM 888  
QY 597 GCGCGTCTTGTAGACCTACATGATGCAATCTTAAGAGAGGGATCTGACACATTTCTA 656  
DB GGMATVGMGGVGMGGTGTGKTMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGM 828  
QY 657 AGGCATCCATATAGGCATTTGGGCGCTAAGTCGCAATTGAAGAGAGATAAGGGGGTGTGAA 716  
DB NGKTMHGKNNMMVMVKTMMVGAAGAAADAGGAAAGAAAGAAAGAAAGAAAGAKAGKHMN 768  
QY 717 AGTGGTGTGCAAAAGAGAGTTCGATTCGCTATACCGCCGCTAAGAGGGTGGGCTAGCAG 776  
DB KMGKTMGTMMGTGKMGKMGKTMGTMMGTMMGTMMGTMMGTMMGTMMGTMMGTMM 708  
QY 777 CTGCTGCGAGCTGGAATACGTCATCTGCTTAGTATGTCACCTAACTCTCAGCAGATG 836  
DB KMGKTMGTMMGTGKMGKMGKTMGTMMGTMMGTMMGTMMGTMMGTMMGTMMGTMM 648  
QY 837 CAAATGCTGATTCGGTTAAATGGGATGATGTTAGGTGTCGGAATACGCTGTTGGATACACGTTG 896  
DB MNNMKKTKKKNN 589  
QY 897 AGTTAAAGGGAAGCTGAAAGCTGAACCTGTCAAGAAATACGCTGTTGGATACACGTTG 956  
DB MNN 529  
QY 957 ATAAACCAATTCAGTCTCAAGGGTGTCTCTGATATGCTGGAGCTTCCCTGTCGCAATGTG 1016  
DB MAMMM 469  
QY 1017 GGGTAACATATTCAT 1031  
DB MGGMMNN 454

RESULT 3

BH285801  
 LOCUS  
 DEFINITION CH230-118D10.TVB CHORI-230 Segment 1 Rattus norvegicus genomic clone CH230-118D10, genomic survey sequence.  
 ACCESSION BH285801  
 VERSION BH285801  
 KEYWORDS GSS.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 REFERENCE 1 (bases 1 to 557)  
 AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shwartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, P., de Jong, P., and Fraser, C.M.  
 TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
 JOURNAL Unpublished (1999)  
 COMMENT Other\_GSSs: CH230-118D10.TVB  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library availability, please contact Pieter de Jong ([pdejong@mail.choi.org](mailto:pdejong@mail.choi.org)). Clones may be purchased from BACPAC Resources ([http://www.chori.org/bacpac/orering\\_information.htm](http://www.chori.org/bacpac/orering_information.htm)). BAC end page: [http://www.tigr.org/tldb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/rat/bac_end_intro.html)  
 Place: 118 row: D column: 10  
 Seq primer: T7  
 Class: BAC ends.  
 Location/Qualifiers  
 1. 557  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-118D10"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_lib="CHORI-230 segment 1"  
 /note="vector: PTABAC2.i; Site 1: EcoRI; Site 2: EcoRI; CHORI-230 Rat (BN/SNHsd/MCW) BAC library produced by Pieter de Jong"

Query Match 3.0%; Score 41.6; DB 28; Length 557;  
 Best Local Similarity 49.1%; Pred. No. 0.85;  
 Matches 110; Conservative 0; Mismatches 114; Indels 0; Gaps 0;  
 QY 1155 AGATGGCGGCTTGAAGTAGCTACCGCATCGAACCCCGAAGCGGTTCAAGACATGGCGGT 1214  
 DB 89 AGATGGCTGGAATTAAAGAGATGACAAAGAGAGGGGGAGACATATGAAGAGTCATCAGG 148  
 QY 1215 ACGTAGATACATAGAGTCATAGACATATAAGGAGCTTGAAGAACCATTCATCAATCCTAA 1274  
 DB 149 ACATGTATGCACATATTACTAGTAGCTACGTTCAAGAGGACATCATTTCTATTAAATATTATA 208  
 QY 1275 GGGTCTCTCTTCTTCTGTCATCACATCAAGAAATCATACACTCAAAACCGAGGAACCTTTCT 1334  
 DB 209 CGTTATCATGTTTCTTCTTAAAAAAAATAAAGTTAAAGATACAGAGGAACCTCTATGA 268  
 QY 1335 ATCTTCCTATAGCAATTCCTCAAAACCATCAATCAACCTTACA 1378  
 DB 269 GACTGAACCTAGAGAAATCCCAACCCCTCTAGAGATACATAGCA 312

RESULT 4  
 CNS05QQW/c

CNS05QQW 1031 bp DNA linear GSS 01-SEP-2000  
 Tetraodon nigroviridis genome survey sequence SP6 end of clone 025B03 of library B from Tetraodon nigroviridis, genomic survey sequence.  
 ACCESSION AL349601  
 VERSION AL349601.1  
 KEYWORDS GSS; genome survey sequence.  
 SOURCE Tetraodon nigroviridis  
 ORGANISM Tetraodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.  
 REFERENCE 1  
 AUTHORS Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizes, C., Wincker, C., Brottier, P., Quetier, F., Saurin, W., and Weissenbach, J.  
 TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence  
 JOURNAL Nat. Genet. 25 (2), 235-238 (2000)  
 MEDLINE 20296633  
 PUBMED 10835645  
 REFERENCE 2  
 AUTHORS Roest Crolius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C., Fizes, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A., and Weissenbach, J.  
 TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
 JOURNAL Genome Res. 10 (7), 939-949 (2000)  
 MEDLINE 20359837  
 PUBMED 10899143  
 REFERENCE 3 (bases 1 to 1031)  
 Genoscope.  
 Direct Submission  
 TITLE Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr))  
 JOURNAL - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.  
 FEATURES  
 source  
 1. 1031  
 /organism="Tetraodon nigroviridis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:99883"  
 /clone="025B03"  
 /clone\_lib="B"  
 /note="Genoscope sequence ID : COAB025CA02B1-end : SP6"

Query Match 2.8%; Score 39; DB 29; Length 1031;  
 Best Local Similarity 43.1%; Pred. No. 7.2; Indels 0; Gaps 0;  
 Matches 81; Conservative 0; Mismatches 107; Indels 0; Gaps 0;  
 QY 792 AATAAGTCACCTTGCTTAGGTATGTCCACCTAATGTGCAGCAGATGCAAAATGCTGTTGG 851  
 DB 523 AAAAAAGTTTGTTCGTAGCAGTACTTAAATAATNACNAGAAAANAANGNAATAAGT 464  
 QY 852 TTAAGTGGCAGTGTAGTGTAGTGGCGGAAACACGCTTGTAGATCTAGTTAAAGGAGCT 911  
 DB 463 NTAAATTTGNNNNNNNTTATTCATNAGNANATNTGCTGCTNATATNAGTAGTAA 404  
 QY 912 GAAAGCTGAACCTGTGCAGAAATAAGCTGTGTGAATACAAACGTTGATAACCCAAATTCAGT 971  
 DB 403 NAATNCTACCATTTNAAAACGNNANTNATGNANAACCTNAGTTCAAAATCCNACTCTTT 344  
 QY 972 CGTCAAGG 979  
 DB 343 CCTNAGGG 336

RESULT 5  
 CNS05QQW/c

```

BX381961      1201 bp      mRNA      linear      EST 08-MAY-2003
LOCUS      BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION      clone CSODI072YF05 3-PRIME, mRNA sequence.
ACCESSION      BX381961
VERSION      BX381961.1 GI:30453007
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLES      1 (bases 1 to 1201)
JOURNAL      Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
COMMENT      Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CSODI072CC03NP1.
Location/Qualifiers
FEATURES
source
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI072YF05"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match      2.8%; Score 39; DB 13; Length 1201;
Best Local Similarity 3.5%; Pred.No. 7.8;
Matches 18; Conservative 159; Mismatches 343; Indels 0; Gaps 0;

QY      687 CGGCAATTAAGGAGATAGGGGGGTGTGAAGATGGTGTGTCAAAAGGAGGTCGATGGCT 746
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      385 CKANNKKKMKACVANNNNKKCKMMNNNNKKKKKKKKKKKKKKKKKKKKKKKKKK 444
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      747 ATACGAGCGGCTAAGCAGCGGGGTAGCAGCTGTCTGCAGCTGTGGAATACGTCACCTGC 806
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      445 NNCCNANNMMNNKKKKKKKKNNNNNTKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 504
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      807 TTAGGTATGTCACCTAATGTACAGATGCAATGTCTGATTGGGTTAAATGGGCATGT 866
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      505 CANNKKCKMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 564
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      867 AGTGTAGGTGCGGAAACACGTTTAGATCTTAGTTAAAGGGGAGCTGAAAGCTGAACCTGT 926
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      565 NNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 624
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      927 CAGAAATAAGCCTGTGGAATAACAGCTGTGATACCCCAATTCAGTCGTCAGGGGTGCT 986
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      625 KMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 684
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      987 GATATGCTGGAGCTTCCCTGTGCGATGTGTGGGGTAACCTATTTCATAGTGGGGCAGAAATGC 1046
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      685 KNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 744
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      1047 AACTCTATTTCATTTGAATTAACCTATTTCGCGGTAGGATGTTCTCAATGTTCTTCTCGC 1106
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      745 NCNKKNNKNTMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 804
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      1107 TGTCACTTACACATCATGGGGGTCAACACAGCTATACAGCTTCATAGAGAGTGCAGCAT 1166
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      805 KKKKKKMKGMVCKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 864
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY      1167 TGAAGTAGTACCGCATCGAACCAGCGAGCGGTTCAAGAC 1206
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      865 NKAMKKKDNMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 904
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
CC264185      1205 bp      DNA      linear      GSS 13-MAY-2003
LOCUS      CH261-57E7_Sp6.1 CH261 Gallus gallus genomic clone CH261-57E7,
DEFINITION      genomic survey sequence.
ACCESSION      CC264185
VERSION      CC264185
KEYWORDS      GSS.
SOURCE      Gallus gallus (chicken)
ORGANISM      Gallus gallus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
TITLES      1 (bases 1 to 1205)
JOURNAL      Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
COMMENT      Warren, W., Graves, T., Mardis, E. and Wilson, R.
Unpublished (2003)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: Sp6 ATTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 50
High quality sequence stop: 729.
Location/Qualifiers
FEATURES
source
1..1205
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strains="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-57E7"
/sex="female"
/cell_line="UCD001, inbred 256"
/notes="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: http://www.chori.org/bacpac"
ORIGIN
Query Match      2.8%; Score 38.8; DB 28; Length 1205;
Best Local Similarity 48.2%; Pred.No. 9;
Matches 109; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY      823 AATGTCAGCAGATGCAATGCTGATGGGTAAATGGGCATGTAGTGTGCGGAAA 882
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      845 AATCTCCACTGACAGATTTGATTTGGGATGAAGGAGGCGCTTGTGGGGCAGGAA 904
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      883 ACACGTTTATAGTCTAGTTAAAGGGAGCTGAAGCTGAACCTGTTCAGAAATAAGCTGTT 942
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      905 AAAAGTTAACCCCAATTCATAAATGAACCCCTCAGTAAACCTTAAAGATCGAACCAAGA 964
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      943 GGAATACAGCTGTATACCCATTCAGTCTGTCAGGGTGTCTCTGATATGCTGAGCTTC 1002
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      965 AGAATAAAGGGGGTCTATCAACCTTGGTTATGAAGGGGTATCCCATTAATGTTTCCITTA 1024
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      1003 CTGTGCGCATTTGGGGTAACTATTTCATAGTGGGGCAGAAATGCAA 1048
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1025 ACCTTCATTCGGGCAACAAAATTCATGCGAGGGGGAGAAAATAA 1070
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
BZ246183/c
LOCUS      BZ246183
DEFINITION      CH230-374A18.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-374A18, genomic survey sequence.

```



```

ACCESSION   B2246183
VERSION     B2246183.1  GI:23906447
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

REFERENCE   1 (bases 1 to 673)
AUTHORS    Zhao S., Shetty J., Shatsman S., Teagay G., Geer K.,
            Shwartsbeyn A., Gebregeorgis E., Overton L., Russell D., Chen D.,
            Riggs F., de Jong P. and Fraser C.M.
TITLE      Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL    Unpublished (1999)
COMMENT    Other GSSs: CH230-374A18.TU
            Contact: Shaving Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.html). For BAC library
            availability, please contact Pieter de Jong (pjejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or erling/information.htm). BAC end
            page: http://www.tigr.org/tcdb/bac_end/rat/bac_end_intro.html
            Plate: 374 row: A column: 18
            Seq primer: T7
            Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..673
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SSHsd/MCW"
                     /db_xref="taxon:10116"
                     /clones="CH230-374A18"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
                     CHORI-230 Rat (BN/SSHsd/MCW) BAC library produced by
                     Pieter de Jong"

ORIGIN
Query Match       2.8%; Score 38.4; DB 28; Length 673;
Best Local Similarity 51.1%; Pred. No. 8.6;
Matches 90; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy 1100 TCTCGCTGTCTACATCATCTGGGGTCAACACGATATACAGCTTCATAGAGT 1159
      |||||
Db 506 TCTGTGTTAAATTAATAAGAAATGCTATTTATGTATATATATATATACACAGAGA 547
      |||||

Qy 1160 GCGGATTTGAAGTAGCTACCCATCGAACCCGGAAGCGTTCAAGACATGGCGTAGCTA 1219
      |||||
Db 546 GAGACACAGACAGACAGACACAGACACAGACAGACAGACAGACAGACACAGACAGA 487

Qy 1220 GATACATAGATCATAGAAATAAAGAGGCTTGAAGACCATTCAAATCCTAAG 1275
      |||||
Db 486 CACAGATACACAGACACAGACAGAGGGAGATTCTCGAAGCTTCTCAAGAGAAAG 431

RESULT 8
AQQ18895/c
LOCUS       AQQ18895
DEFINITION  CIT-HSP-2302116.TF CIT-HSP Homo sapiens genomic clone 2302116,
            genomic survey sequence.
ACCESSION   AQQ18895
VERSION     AQQ18895.1  GI:3197631
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens

Query Match       2.8%; Score 38.2; DB 28; Length 479;
Best Local Similarity 50.3%; Pred. No. 8.2;
Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 31 GTGTCCAAAACCTGTGTAGTGACTACGAATGAGGAAAGAAACGGTGTGTGTGGCAGCT 90
Db 460 GTGTACAAAATGCTAAAGAAATTTTGAAGAAAGAAAGAAATAGTTCAATGGTGGCAGGG 401
Qy 91 GAGACTGAGAGAGGAGCCAAAGATATTCACATGCGATACCGTTGCATCATGCTTGT 150
      |||||
Db 400 AGTGAAGAAAGTAGAGGGTATAGATAATATGAATGGTTATATCGGTGTGGC 341
Qy 151 TCAAGAGACGACCGTTCATCTCTGCTGTTCCTCTCTTCCTTGTGTACAGATCAAGTATC 210
      |||||
Db 340 TAGGCAATGGACATTTGATTGACGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAT 281
Qy 211 GGATGAC 217
      |||||
Db 280 GGAATAC 274

RESULT 9
CNS04NPT/c
LOCUS       CNS04NPT
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
            123C09 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL299018
VERSION     AL299018.1  GI:8038159
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
ORGANISM    Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Teleostei; Euteleostei; Neoteleostei;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodontidae; Tetraodon.
            Roest Crolius H., Jaillon O., Dasilva C., Bouneau L., Fisher C.,
            Adams M.D., Rounsley S.D., Zhao S., Field C.E., Bass S., Linher K.,
            Golden K., Berry K., Granger D., Suh E., Wible C., Shizuya H.,
            Simon M. and Venter J.C.
            Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
            Unpublished (1998)
            Other GSSs: CIT-HSP-2302116.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tcdb/hungen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..479
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2302116"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
                     HindIII"

ORIGIN
Query Match       2.8%; Score 38.2; DB 28; Length 479;
Best Local Similarity 50.3%; Pred. No. 8.2;
Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 31 GTGTCCAAAACCTGTGTAGTGACTACGAATGAGGAAAGAAACGGTGTGTGTGGCAGCT 90
Db 460 GTGTACAAAATGCTAAAGAAATTTTGAAGAAAGAAAGAAATAGTTCAATGGTGGCAGGG 401
Qy 91 GAGACTGAGAGAGGAGCCAAAGATATTCACATGCGATACCGTTGCATCATGCTTGT 150
      |||||
Db 400 AGTGAAGAAAGTAGAGGGTATAGATAATATGAATGGTTATATCGGTGTGGC 341
Qy 151 TCAAGAGACGACCGTTCATCTCTGCTGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAT 210
      |||||
Db 340 TAGGCAATGGACATTTGATTGACGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAT 281
Qy 211 GGATGAC 217
      |||||
Db 280 GGAATAC 274

```

Bernot, A., Pizanes, C., Wincker, P., Brottier, P., Quetier, F., Saurin, M. and Weissenbach, J.  
 Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence  
 Nat. Genet. 25 (2), 235-238 (2000)  
 20296633  
 10835645  
 2  
 AUTHORS  
 Roest Crolius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C., Pizanes, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, M., Bernot, A. and Weissenbach, J.  
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
 Genome Res. 10 (7), 939-949 (2000)  
 20359837  
 10899143  
 3 (bases 1 to 923)  
 Genoscope.  
 Direct Submission  
 Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.  
 Location/Qualifiers  
 1. 923  
 /organism="Tetraodon nigroviridis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:99083"  
 /clone="123C09"  
 /clone\_lib="C"  
 /note="Genoscope sequence ID : COBG123AB05SP1-end : pUC-ori"

Query Match 2.8%; Score 38.2; DB 29; Length 923;  
 Best Local Similarity 50.3%; Pred. NO. 12;  
 Matches 85; Conservative 3; Mismatches 81; Indels 0; Gaps 0;

QY 891 AGATCTAGTTAAAGGAGCTGGAAGCTGCAAGCTGCAAGCTGTCAGAAATATGCGTATGGAATACA 950  
 :|||  
 DB 352 RGAGAAAGAGAAAAAANKCCATGCTGTCATATCACAACTATGATGTGGCGTGG 293

QY 951 ACCTTGATACCAATTCATGCTGTCAGAGGTGCTGATATGTCGAGTTCCTGTCGC 1010  
 DB 292 GCCTTAATATCTTACAAATGTTGGGATGGAACGCCGACTTGGCGGCTCTTCGCTTAAT 233

QY 1011 ATGTGCGGTAACTATTTTCATGTGGGCGAGATGCAACTCTATTTTCA 1059  
 DB 232 ATCTCGCTCTCTTTGATTTCCGGTGAACAACAAAGACAATATCA 184

RESULT 10  
 BG185531/c  
 LOCUS  
 DEFINITION  
 RST4479 Atherys RAGE Library Homo sapiens cDNA, mRNA sequence.  
 BG185531  
 VERSION  
 BG185531.1 GI:13707218  
 EST.  
 KEYWORDS  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 482)  
 Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whitington, J., Lerner, L., Cosanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Kikka, A., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J. and Ducar, M.  
 Creation of genome-wide protein expression libraries using random activation of gene expression

TITLE



Contact: Kazuhiro W. Makabe  
Department of Zoology, Graduate School of Science  
Kyoto University  
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
Tel.: 81-75-753-4095  
Fax: 81-75-705-1113  
Email: kwmakabe@es.sci.dian.zool.kyoto-u.ac.jp.

```

/cw_xcels=xcels017123
/cfones=008M08_5"
/dev_stage=Fertilized egg"
/clone lib=Halocynthia roretzi fertilized egg"

```

Query Match	2.7%;	Score 37.6;	DB 9;	Length 688;
Best Local Similarity	54.3%;	Pred. No. 15;		
Matches 76;	Conservative 0;	Mismatches 64;	Indels 0;	Gaps 0;
1056	TTCAATTGAATCTAAACTATTCTGGTATGGAGTTCTCAATGTCCTTTCTCGCTGTCTACTTA	1115		
	20	TGCAAGTGAACCCACAGGTGTTGAGAGTAGGAACTCGCAAAAGCCCTCTCGCAATGATTC	79	
1116	CACACATCATGGGGGTCAACAACGTATACAGCTTCATAGAGAGTGGCGCATTTGAAGTAGC	1175		
	80	GACACATGGGTAGTAAATGAACTCGAAAAAAGCTCAACAGCAACTTGGATATAGAAATAGT	139	
1176	TACCGCATCGAACCCGGGAG	1195		
	140	GCCTATGTCAACCAATTGGAG	159	

## RESULT 15

LOCUS	AZ960759	AZ960758	598 bp	DNA	linear	GSS 27-APR-2001
DEFINITION	2M0228L20R Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0228L20 R, genomic survey sequence.					

NO. 15384

## KEYWORDS

## SOURCE

ORGANISM

## REFERENCE

## AUTHORS

**ITEM 8**

STTTT

FORTRAN.

COMMENT	REPLY
<p>1. The authors state that the "mean age of the patients was 67 years." However, the mean age of the patients was 67 years, not 67 years.</p>	<p>The authors state that the "mean age of the patients was 67 years." However, the mean age of the patients was 67 years, not 67 years.</p>

## TABLE 1

RESIN.T 14

AV384223

LOCUS

### DEFINITION

**ACCESSION**

**VERSION**

## KEYWORDS

**SOURCE**

## ORGANISMS

**\*\*\*\*\***

3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527

AUTHORITY  
 TTTT.R

JOURNAL

**TABLE 1**

10



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 08:08:35 ; Search time 421.164 Seconds

(without alignments)  
13899.608 Million cell updates/sec

Title: US-10-070-386-1

Perfect score: 1378

Sequence: 1 gtcgacgtgggtggatgat.....accatcaatcaacacaa 1378

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04.\*

1: Geneseqn1980s.\*

2: Geneseqn1990s.\*

3: Geneseqn2000s.\*

4: Geneseqn2001as.\*

5: Geneseqn2001bs.\*

6: Geneseqn2002s.\*

7: Geneseqn2003as.\*

8: Geneseqn2003bs.\*

9: Geneseqn2003cs.\*

10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1378	100.0	1378	5	Aaf79834 Mycelia s
2	48.2	3.5	2000	7	Ada71938 Rice gene
C 3	40	2.9	6197	6	Abn80257 Human che
C 4	37.2	2.7	2000	7	Ada71938 Rice gene
C 5	36.2	2.6	2672	6	Aad29004 Bugula ne
C 6	36.2	2.6	4590	5	Aab24065 Yeast AOD
C 7	35.4	2.6	993	4	Aaf83969 Human SER
C 8	35.4	2.6	993	9	Adc29362 Human cDN
C 9	35	2.5	5971	6	Ab133237 Human imm
C 10	35	2.5	207433	5	Ab372040 Gene 216
C 11	35	2.5	207433	7	Abx74891 BAC1098L2
C 12	34.6	2.5	5059	2	Abx84332 Stealth v
C 13	34.4	2.5	2799	7	Abv74497 Human typ
C 14	34.4	2.5	8805	6	Abk40016 Human che
C 15	34.4	2.5	32874	8	Ada02648 Human TBX
C 16	34.4	2.5	32874	9	Adc72386 Human TBX
C 17	34.2	2.5	1500	9	Adc92243 Human che
C 18	34	2.5	6283	6	Abk39991 Human che
C 19	34	2.5	6283	6	Ab132834 Human imm
C 20	33.4	2.4	403	4	Aai36496 Probe #50
C 21	33.4	2.4	403	4	Abc26537 Probe #51
C 22	33.4	2.4	1781	2	Aav35130 Mouse WRN
C 23	33.4	2.4	4206	2	Aav35115 Mouse WRN

24	33.4	2.4	4792	2	Aax83004	Aax83004 Mouse WRN
25	33.4	2.4	5058	2	Aav35114	Aav35114 Mouse WRN
26	33.4	2.4	6476	2	Aax24302	Aax24302 Murine mw
27	33.4	2.4	7189	4	AAS30652	AAS30652 DNA encod
28	33.4	2.4	7189	4	AAS28725	AAS28725 Genomic s
29	33.4	2.4	7189	7	ACA03415	ACA03415 DNA encod
30	33.4	2.4	7189	8	ADB96763	ADB96763 Novel lun
31	33.4	2.4	16442	2	AAX83006	Aax83006 Partial m
32	33.2	2.4	891	4	AAK83873	AAK83873 Human imm
33	33.2	2.4	892	4	AAK82975	AAK82975 Human imm
C 34	33.2	2.4	6394	6	ABK31374	ABK31374 Signal tr
C 35	33.2	2.4	6394	6	AAAS61278	AAAS61278 Human gen
C 36	33.2	2.4	14253	6	ABL33495	ABL33495 Human imm
C 37	33.2	2.4	22214	6	ABL01222	ABL01222 Human AKR
C 38	33.2	2.4	22214	6	ABL01105	ABL01105 Human AKR
C 39	33.2	2.4	34548	6	ABL70603	ABL70603 Chemocall
C 40	33	2.4	429	6	ABL64333	ABL64333 Stomach c
C 41	33	2.4	429	6	ABL61677	ABL61677 Colon ade
C 42	33	2.4	478	3	AAAC36866	AAAC36866 Arabidops
C 43	33	2.4	1189	4	AAK83871	AAK83871 Human imm
C 44	33	2.4	2207	4	ABL24548	ABL24548 Drosophil
45	32.8	2.4	414	8	ADA29990	ADA29990 DNA encod

## ALIGNMENTS

### RESULT 1

AAf79834  
ID AAF79834 standard; DNA; 1378 BP.  
AC AAF79834;  
XX  
DT 30-MAY-2001 (first entry)  
XX  
DB Mycelia sterilia promoter sequence.  
XX  
KW Promoter; terminator; regulatory region; filamentous fungus;  
protein production; ds.  
XX  
OS Mycelia sterilia.  
XX  
FN WO200118219-A1.  
XX  
PD 15-MAR-2001.  
XX  
PF 07-SEP-2000; 2000WO-JP006104.  
XX  
PR 07-SEP-1999; 99JP-00252851.  
XX  
PA (MEIJ ) MEIJI SEIKA KAISHA LTD.  
XX  
PI Watanabe M, Murakami T;  
XX  
DR WPI; 2001-235202/24.  
XX  
PT New promoter and terminator functioning synchronously, useful for  
regulating expression of endogenous gene in filamentous fungi,  
particularly transformed Mycelia sterilia capable of producing high  
yields of target protein/substance.  
XX  
PS Claim 2; Page 18-20; 26pp; Japanese.  
XX  
CC The present invention provides the sequences of a promoter and terminator  
from Mycelia sterilia which can be used to regulate the expression of an  
endogenous gene in filamentous fungi, particularly transformed Mycelia  
sterilia, to produce useful target proteins or other substances. The  
present sequence is the promoter of the invention  
XX  
SQ Sequence 1378 BP; 396 A; 289 C; 340 G; 353 T; 0 U; 0 Other;

Query Match 100.0%; Score 1378; DB 5; Length 1378;  
Best Local Similarity 100.0%; Pred. No. 0;



	Matches	1378:	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	GTGACGTGGGTGGTGTGATATCATGGCTGTGTGTGTCCTAAACCTGTGTGTAGTGA	60							
Db	1	GTGACGTGGGTGGTGTGATATCATGGCTGTGTGTGTCCTAAACCTGTGTGTAGTGA	60							
Qy	61	GAGGAAAGAAACCGTGGTGTGTGGCAGCTGAAGACTGAAGAGGAGCCAAAGATAATTC	120							
Db	61	GAGGAAAGAAACCGTGGTGTGTGGCAGCTGAAGACTGAAGAGGAGCCAAAGATAATTC	120							
Qy	121	ACAAATGCATACGGTTGCATCAATGCTTGTTCAGAGAGCAGCTTGCGATCTPACCTGGTGT	180							
Db	121	ACAAATGCATACGGTTGCATCAATGCTTGTTCAGAGAGCAGCTTGCGATCTPACCTGGTGT	180							
Qy	181	TCCCCTCTTGGTTGTACAAGATCAAGATTCGGATGACACCCACCCCGCAACGGAAATCTGT	240							
Db	181	TCCCCTCTTGGTTGTACAAGATCAAGATTCGGATGACACCCACCCCGCAACGGAAATCTGT	240							
Qy	241	GAGTTCAAAGAGGGTGTCTCAGGCATTTAGGTATAGATAGGCATAGGGTTTCAAGCTAA	300							
Db	241	GAGTTCAAAGAGGGTGTCTCAGGCATTTAGGTATAGATAGGCATAGGGTTTCAAGCTAA	300							
Qy	301	GCTGAAAGCTGATTTACGAGACATGAGACACAGAAATAACAACGGTGTGTATGCGTTC	360							
Db	301	GCTGAAAGCTGATTTACGAGACATGAGACACAGAAATAACAACGGTGTGTATGCGTTC	360							
Qy	361	GCTTACTAAAGTGTATCCAGAGACACAGCCCGAAGAAACCGATGCTGTCTGAGGGG	420							
Db	361	GCTTACTAAAGTGTATCCAGAGACACAGCCCGAAGAAACCGATGCTGTCTGAGGGG	420							
Qy	421	TTCCCTTAGAGTCTACATGCTGTAAACGGTGTGATAGAGAAACATCAAAATGGCCAAATCA	480							
Db	421	TTCCCTTAGAGTCTACATGCTGTAAACGGTGTGATAGAGAAACATCAAAATGGCCAAATCA	480							
Qy	481	AGTATACCTGACCGTACATCGCTTTCCTCGGATCTTGCCTAAATAATATGTGCTCTGC	540							
Db	481	AGTATACCTGACCGTACATCGCTTTCCTCGGATCTTGCCTAAATAATATGTGCTCTGC	540							
Qy	541	GAACTGTGCGTACTGCTTCTGCTACTAACTGTCTTCCGTTGAAGTCTCTAGCACAGCCCG	600							
Db	541	GAACTGTGCGTACTGCTTCTGCTACTAACTGTCTTCCGTTGAAGTCTCTAGCACAGCCCG	600							
Qy	601	CGTTGTAGACCTACATGATGTCACATCTTAAAGCAGGGATCTGAGACATTTTCTAAGGC	660							
Db	601	CGTTGTAGACCTACATGATGTCACATCTTAAAGCAGGGATCTGAGACATTTTCTAAGGC	660							
Qy	661	ATCCATATAGGCATTGGCGCTAAGTCGGCATTTGAAGGAGATAGGGGGGTGTGAAATG	720							
Db	661	ATCCATATAGGCATTGGCGCTAAGTCGGCATTTGAAGGAGATAGGGGGGTGTGAAATG	720							
Qy	721	GTGTGTCAAAGAGGCTCGATTGGCTATACACGCGCTAAGCAGGTTGGGCTAGCAGCTGT	780							
Db	721	GTGTGTCAAAGAGGCTCGATTGGCTATACACGCGCTAAGCAGGTTGGGCTAGCAGCTGT	780							
Qy	781	CTGCGCTGTGAATTAAGTCACTTGTCTAGGTATGTCCACTTAATGTCAAGAGATGCAAA	840							
Db	781	CTGCGCTGTGAATTAAGTCACTTGTCTAGGTATGTCCACTTAATGTCAAGAGATGCAAA	840							
Qy	841	TGCTGATTGGGTTAAATATGGCATGTAGTGTAGGTGCCGAAAACACGTTTAGATCTAGTT	900							
Db	841	TGCTGATTGGGTTAAATATGGCATGTAGTGTAGGTGCCGAAAACACGTTTAGATCTAGTT	900							
Qy	901	AAAGGAGAGCTGAAAACCTGTGAGAAATTAAGCTGTTGGATATCAACGTTTGATAA	960							
Db	901	AAAGGAGAGCTGAAAACCTGTGAGAAATTAAGCTGTTGGATATCAACGTTTGATAA	960							
Qy	961	CCCAATTACGTGTCAAGGGTGTCTGATATGCTGGAGCTTCCCTGTGCGATTTGGGGT	1020							
Db	961	CCCAATTACGTGTCAAGGGTGTCTGATATGCTGGAGCTTCCCTGTGCGATTTGGGGT	1020							
Qy	1021	AACATTTTCATGTGGGGCAGAAATGCAACTCTATTTTCAATGAAATCAAACTATTCTCG	1080							
Db	1021	AACATTTTCATGTGGGGCAGAAATGCAACTCTATTTTCAATGAAATCAAACTATTCTCG	1080							

## RESULT 2

RESULT 2  
ADA71938  
ID ADA71938 standard; DNA; 2000 BP.  
XX  
XX AC  
XX ADA71938;  
XX  
XX 20-NOV-2003 (first entry)  
XX  
XX Rice gene, SEQ ID 5263.  
XX  
XX Plant; bacterial infection; fungal infection; viral infection; rice;  
XX  
XX gene; db.  
XX  
XX Oryza sativa.  
XX  
XX WO2003000898-A1.  
XX  
XX 03-JAN-2003.  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
XX Katagiri P, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175290/17.  
XX  
XX

XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection or conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.

XX  
XX  
PS Claim 27; SEQ ID NO 5263; 899pp; English.

XX  
XX  
CC The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.

Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;  
Query Match 3.5%; Score 48.2; DB 7; Length 2000;

[illegible]

XX	Rice gene, SEQ ID 5263.
XX	DE Plant; bacterial infection; fungal infection; viral infection; rice;
XX	XW gene; ds.
XX	OS Oryza sativa.
XX	PN W02003060899-Al.
XX	XX 03-JAN-2003.
XX	PD 22-JUN-2001; 2001WO-IB001105.
XX	PF 22-JUN-2001; 2001WO-IB001105.
XX	PR (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX	PA Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX	PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX	PI WPI; 2003-175290/17.
XX	DR Identifying at least one gene involved in plant resistance or response to
XX	PT pathogenic infection for conferring resistance or tolerance to a plant to
XX	PT bacterial, fungal or viral infection by determining or detecting plant
XX	PT gene expression.
XX	PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX	The present invention relates to a method (M1) for identifying genes
XX	CC involved in plant resistance or response to pathogenic infection. M1
XX	CC comprises identifying a gene whose expression is significantly altered in
XX	CC the incompatible interaction of plant gene expression relative to
XX	CC expression of the gene in an uninfected plant, in a mutant plant that
XX	CC does not express a gene associated with response to pathogenic infection,
XX	CC or in a corresponding incompatible or compatible interaction. (M1) is
XX	CC useful for conferring resistance to resistance or tolerance to a plant to
XX	CC bacterial, fungal or viral infection. The present sequence was used to
XX	CC illustrate the invention.
XX	Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
SQ	Query Match 2.7%; Score 37.2; DB 7; Length 2000;
	Best Local Similarity 10.2%; Pred. No. 0.97;
	Matches 42; Conservative 190; Mismatches 178; Indels 2; Gaps 1
Qy	362 CTTACTAAGTCATATCCAAGCACAACACGCGCGAAGAACCAGTGCTGTCTGAGGGT 421
Dd	557 KKKKCTWRCMCRMGTYMTYTSRSRMYYGRVKGARTYSKRRTMYTKIRKYCYWYYGYX 498
Qy	422 TCCTTTAGAGTCATCATGTAACGGTCGCATGATAGAAACATCAATCGCCAATCAAGTFA 481
Dd	497 XKCSYMRVYGCKACKCCYAMC--WKAAYSGMWYRKYKSWKMSYTKYWSMWYKCC 440
Qy	482 GTATACCTGACGTCATCGCTTTCTTCOGGATCTGCCCTAAATAATATATGCTGTCG 541
Dd	439 RSMKYGAKGCGCKRWYTCVGYMKWTYPYGSYKRCYKYMRYMYKGMWNTMYYSAYS 380
Qy	542 AACTGTCGCTACTGCTTCGTACTACTGTCCTCCGTTGAAGTCCTTAGCACAGCGCGC 601
Dd	379 SMMTWYYYAKWYKWYKERGMTSMYSGSKYKXYCTWMCYMKCR CYRMRKWBKTKYS 320
Qy	602 GTTTGTAGACCTACATGATGCCACATCTTAAAGCAGGGATCTGAGACATTTCTTAAGCA 661
Dd	319 KRCYCWRVATCYCCCYRKGMYSRWMRTAGSKWRMSRCSRYSYKMYKDMYKKS 260
Qy	662 TCCATATAGGCATTGGGGCGCTAAGTCGSCATTGAAGGAGATAAGGGGGGTGTCAAAGTG 721
Dd	259 YNMSGARSSCTWSRSAXKFTYKYSTSRMRACRMYSA CRYSSTSYSCGSYCG 200
Qy	722 TGTGTCAAAGGAGCTGATGGCTATACCGCGCTTAACGAGGTGGGTAG 773



CC	represents the nucleotide sequence of human SER4														
XX	Sequence	993	BP;	229	A;	305	C;	264	G;	195	T;	0	U;	0	Other;
SQ	Query Match	2.6%	Score	35.4;	DB	4;	Length	993;							
	Best Local Similarity	54.1%;	Pred. NO.	2.6;											
	Matches	72;	Conservative	0;	Mismatches	61;	Indels	0;	Gaps	0;					

22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051

Q2 504 AAATCAGTCGTCAAGGGTGTCCTCGATATGCTGGAGCTTCCCTGTCGCAATGTCGGGTAAC 1023

DD AG:IGTCCCTGGHGGIGTCAIACACIACACCGACGACGCTCGAHHGCCCCIIGAGGCTGGC 389

QY 1024 1A111C1A1AG1GG 1036  
| | | | | | | |

DD CHGCTCAGAGG 376

RESULT 8  
ADE29362/c  
ID ADE29362 standard; cDNA; 993 BP.  
XX

AC  
AC  
XX  
ALJE29362;

DT 29-JAN-2004 (first entry)  
XX

DE Human CDNA encoding serine/threonine kinase SER4.  
XX

Human; ss; gene; serine threonine kinase; SER4;  
blood coagulation disorder: liver disorder: hematoma cell disorder:

hepatocellular cell disorder; cancer; gene therapy.

OS Homo sapiens.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

```

/*tag= a
FT

```

XX  
XXXXXX  
XXXXXX

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

Page No.

Date

[illegible]

ED  
XX  
Z4-AFK-2003.

03-JUL-2001; 2001US-00898837.  
PF  
yy

PR 17-NOV-1999; 99US-0165986P.  
00-PPP-2000; 2000US-0181347D  
00-PPP-2000; 2000US-0181347D

03-APR-2000; 2000US-0194195P.

PR 07-APR-2000; 2000US-0195637P.

PR 03-JUL-2000; 2000US-0215906P.

[illegible]

(SERIAL) SERIAL Y L:  
PA (MACD/) MACDOUGALL J R.

PA (YAJU/) MAJUMDER K,

SPILLER A. A.  
VERNET C.  
BURGESS C. E.  
FERNANDES E. R.  
RASTELLI L.  
HERRMANN J. L.  
SPADENA S. K.  
SHIMKETS R. A.  
TAUPIER R. J.

Gerlach VI, MacDougall JR, Quinn KE, Majumder K, Spytek KA;  
Vernet C, Burgess CE, Fernandes ER, Rastelli L, Herrmann JL;

Gerlach VL, MacDougall JR, Quinn KB, Majumder K, Spytek KA;  
PPI Vernet C, Burgess CE, Fernandes ER, Rastelli L, Herrmann JL;  
PPI

PI Spaderna SK, Shimkets RA, Taupier RJ;  
XX WPI; 2003-615999/58.  
DR P-PSDB; ADE23363.  
XX  
XX New serine/threonine protein-kinase like polypeptides and genes, useful  
PT in gene therapy, diagnosis or prognosis of e.g. diseases of blood  
PT coagulation, obesity; diabetes, or cancers.  
XX  
XX  
PS Claim 9; SEQ ID NO 8; 77pp; English.  
XX  
XX The invention relates to a new isolated polypeptide has a sequence  
CC (designated SER1, SER2, SER3, SER4, SER5 and SER6, respectively) fully defined  
CC in the specification for their mature forms, variants or fragments). Also  
CC included are an isolated nucleic acid molecule (comprising: a nucleic  
CC acid sequence encoding any of the SER polypeptides above; a nucleic acid  
CC fragment encoding at least a portion of the SER polypeptides or their  
CC variants; or their complements), vectors comprising the nucleic acid  
CC molecules, cells comprising the vectors, an antibody that binds  
CC immunospecifically to the SER polypeptide, determining the presence or  
CC amount of the SER polypeptide or nucleic acid in a sample, identifying an  
CC agent that binds to the SER polypeptide, identifying a (potential)  
CC therapeutic agent for use in the treatment of a pathology related to  
CC aberrant expression or aberrant physiological interactions of the SER  
CC polypeptide, modulating the activity of the SER polypeptide,  
CC pharmaceutical compositions comprising a carrier (and the SER  
CC polypeptide, nucleic acid molecule or antibody), kits comprising in one  
CC or more containers any of the compositions above, screening for a  
CC modulator of activity or latency or predisposition to a pathology  
CC associated with the SER polypeptide and determining the presence of or  
CC predisposition to a disease associated with altered levels of the SER  
CC polypeptide or nucleic acid molecule. The SER polypeptide, nucleic acid  
CC or anti-SER antibody is useful for treating or preventing a pathological  
CC state or pathology associated with the SER polypeptide in a subject,  
CC particularly a human. These pathologies include diseases of blood  
CC coagulation, human liver, hepatoma cells or hepatocellular cells,  
CC obesity, diabetes, or cancers. The polypeptide or polynucleotide is also  
CC useful in gene therapy, diagnostics, prognosis, or research. The present  
XX sequence encodes SER4.  
XX  
SQ Sequence 993 BP; 229 A; 305 C; 264 G; 195 T; 0 U; 0 Other;  
Query Match 2.6%; Score 35.4; DB 9; Length 993;  
Best Local Similarity 54.1%; Pred. No. 2.6;  
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
QY 904 GCGAAGCTGAAAGCTGACCTGTCAGAAATACCTGTGGATACAAAGTTGATTAACCC 963  
Db 708 GCGAAGCTGAGGCTGTACGGTCGATAGAGGGCTGCTGGCGAGGCGATAGATTCTC 649  
QY 964 AATTGAGTCTCAAGGGTGTCTGTATGCTGGAGCTTCCTGTCCATTTGGGGTAAC 1023  
Db 648 AGTTGTCCCTGGAGGCTGTATACACTCACCAGGCTCGAAGCCCTCTTGAGGCTGGC 589  
QY 1024 TATTTATAGTGG 1036  
Db 588 CAGCTCGTAGAGG 576  
RESULT 9  
ID ABL33237/c  
XX ABL33237 standard; DNA; 6971 BP.  
AC ABL33237;  
XX  
XX 26-MAR-2002 (first entry)  
DT  
DE Human immune system associated gene SEQ ID NO: 1210.  
XX Human; immune system disease; cytosine methylation; antiaesthatic;  
KW antiaesthatic; antiaesthatic; antiaesthatic; cytosine; nontropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;

KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
KW ds.  
XX Homo sapiens.  
OS  
XX WO2000200928-A2.  
PN  
XX  
PD 03-JAN-2002.  
XX  
XX 02-JUL-2001; 2001WO-BP007537.  
PF  
XX 30-JUN-2000; 2000DE-01032529.  
PR  
XX 01-SEP-2000; 2000DE-01043826.  
XX  
XX (EPIC-) EPICENOMICS AG.  
PA  
XX Olek A, Piepenbrock C, Berlin K,  
PI WPI; 2002-130909/17.  
DR  
XX Nucleic acid comprising fragment of chemically modified gene, useful for  
PT diagnosis and treatment of diseases associated with abnormal cytosine  
PT methylation.  
XX  
PS Claim 1; SEQ ID NO 1210; 32pp + Sequence Listing; German.  
XX  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention  
XX  
SQ Sequence 6971 BP; 1665 A; 242 C; 1831 G; 3232 T; 0 U; 1 Other;  
Query Match 2.5%; Score 35; DB 6; Length 6971;  
Best Local Similarity 53.2%; Pred. No. 10;  
Matches 74; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
QY 1238 AACATAAAGGAGCTTGAAGACCAATTCATAATCTTAAGGCTCTCTTTCTGATCA 1297  
Db 2068 AAAAAAATAAACTATAAAACCTTAAACCTCCAAAGCTCTCTTCTTCTTAAACA 2009  
QY 1298 CATCAAGATCATACACTCAAAACCAAGAACTTTTCTATCTTCTTATAGCAATTCCTCA 1357  
Db 2008 AAACCTATACTATACAAAAATCCCAAAAAAACTCTAAAGCTCCCTCACTCCACCCCA 1949  
QY 1358 AACCCATCAATCACTAA 1376  
Db 1948 CAACCATCCCCCACTAAA 1930  
RESULT 10  
ID ABZ72040/c  
XX ABZ72040 standard; DNA; 207433 BP.  
AC ABZ72040;  
XX  
XX 03-APR-2003 (first entry)  
DT  
DE Gene 216 H194BAC1098L22 nucleotide sequence SEQ ID NO 5.  
XX Human; Gene 216; chromosome 20p13-p12; antiaesthatic; anorectic;  
KW antiinflammatory; gastrointestinal; gene therapy; vaccine; asthma;  
KW obesity; inflammatory bowel disease; promoter; gene; ss.  
OS Homo sapiens.  
XX  
XX WO200178894-A2.  
PN



XX PD 25-OCT-2001.  
XX PF 13-APR-2001; 2001WO-US012245.  
XX PR 13-APR-2000; 2000US-00548797.  
XX PA (GENO-) GENOME THERAPEUTICS CORP.  
XX PI Keith T;  
XX DR WPI; 2001-639428/73.  
XX DR P-PSDB; ABR00926.  
XX PT isolated genes (Gene 216) from human chromosome 20p13-p12 and the  
XX PT proteins they encode, useful for the prevention, diagnosis and treatment  
XX PT of asthma, obesity and inflammatory bowel disease.  
XX PS Example 4; Fig 7; 520pp; English.  
XX CC The invention relates to isolated genes (Gene 216) from human chromosome  
XX CC 20p13-p12 and the proteins they encode. The nucleic acids and proteins  
XX CC may be used in the prevention, diagnosis and treatment of diseases  
XX CC associated with inappropriate Gene 216 expression. For example, the  
XX CC nucleic acids (for vectors) and proteins may be used to treat disorders  
XX CC associated with decreased expression by rectifying mutations or deletions  
XX CC in a patient's genome that affect the activity of Gene 216 by expressing  
XX CC inactive proteins or to supplement the patients own production of Gene  
XX CC 216 proteins. Additionally, the nucleic acids may be used to produce the  
XX CC secreted Gene 216 protein, by inserting the nucleic acids into a host  
XX CC cell and culturing the cell to express the protein. The nucleic acids and  
XX CC complementary sequences may also be used as DNA probes in diagnostic  
XX CC assays to detect and quantitate the presence of similar nucleic acid  
XX CC sequences in samples and therefore which patients may be in need of  
XX CC restorative therapy. The Gene 216 protein may also be used as antigens in  
XX CC the production of antibodies against Gene 216 and in assays to identify  
XX CC modulators of Gene 216 expression and activity. The anti-Gene 216  
XX CC antibodies and antagonists may also be used to down regulate expression  
XX CC and activity. The anti-Gene 216 antibodies may also be used as diagnostic  
XX CC agents for detecting the presence of Gene 216 proteins in samples (e.g.  
XX CC by enzyme linked immunosorbant assay or ELISA). Disorders that may be  
XX CC prevented, diagnosed and/or treated by the above methods include, for  
XX CC example asthma, obesity and inflammatory bowel disease. The present  
XX CC sequence is that of the Gene 216 genomic nucleic acid sequence, promoter  
XX CC or enhancer  
XX SQ Sequence 207433 BP; 52775 A; 51290 C; 51698 G; 51670 T; 0 U; 0 Other;  
Query Match 2.5%; Score 35; DB 5; Length 207433;  
Best Local Similarity 55.3%; Pred. No. 66;  
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 1210 GGCCTAGCTAGATACATAGAGTCTATAGAAACATATAAGAGGCTTGAAGAACCAATTCAAAT 1269  
Db 1335 GTCTTACTCTTCTTCTTAAACACATACACAAATCCAGGGCTAGTAGGACCTTGAAT 1276  
QY 1270 CCTAAGGGTCTCTTCTTCTTGTGATCAGATCAAGAAATCATACACTCAAAACCGGAATC 1329  
Db 1275 ATACAGAACATCTCTTCTTAAAGAAACCTCAGGGATATAAAATCCAAATAGAAACCC 1216  
QY 1330 TTT 1332  
Db 1215 ATT 1213  
RESULT 11  
ABX74891/C  
ID ABX74891 standard; DNA; 207433 BP.  
AC ABX74891;  
XX 07-APR-2003 (first entry)  
XX

DE BAC1098L22 DNA sequence.  
XX Gene 216; antiasthmatic; antiinflammatory; ss; anorectic;  
XX chromosome 20p13-p12; single nucleotide polymorphism; SNP; gene therapy;  
XX respiratory disease; asthma; obesity; bronchial hyper-responsiveness;  
XX chronic obstructive pulmonary disease;  
XX adult respiratory distress syndrome; inflammatory bowel syndrome.  
XX OS Synthetic.  
XX PN WQ200283077-A2.  
XX PD 24-OCT-2002.  
XX PF 15-APR-2002; 2002WO-US012063.  
XX PR 13-APR-2001; 2001US-00834597.  
XX PR 13-APR-2001; 2001WO-US012245.  
XX PA (SCHE ) SCHERING CORP.  
XX PA (GENO-) GENOME THERAPEUTICS CORP.  
XX PI Keith T, Little RD, Van Berdewegh P, Dupuis J, Del Mastro RG;  
XX Simon J, Allen K, Pandit S;  
XX WPI; 2003-092960/08.  
XX DR New isolated gene 216 nucleic acids, useful for diagnosing, preventing or  
XX PT treating a disorder, such as asthma, bronchial hyper-responsiveness,  
XX PT chronic obstructive pulmonary disease, obesity or inflammatory bowel  
XX PT syndrome.  
XX PS Example 6; Fig 7; 650pp; English.  
XX CC This invention relates to a novel isolated nucleic acid, gene 216,  
XX CC identified from human chromosome 20p13-p12. The invention also discloses  
XX CC regions of the 216 gene that contain single nucleotide polymorphisms  
XX CC (SNP's) which may be used as markers for disease susceptibility or  
XX CC severity. The nucleotides of the invention may have antiasthmatic,  
XX CC antiinflammatory or anorectic activities and may be used in gene therapy.  
XX CC The nucleic acids, antibodies or its fragments are useful for diagnosing,  
XX CC preventing or treating a disorder, such as respiratory diseases (e.g.  
XX CC asthma, bronchial hyper-responsiveness, chronic obstructive pulmonary  
XX CC disease or adult respiratory distress syndrome), obesity, or inflammatory  
XX CC bowel syndrome. The nucleic acids are also useful for identifying  
XX CC increased susceptibility of a subject to the disorders mentioned. The  
XX CC nucleic acids can also be used as primers and templates for the  
XX CC recombinant production of disorder-associated peptides or polypeptides,  
XX CC for chromosome and gene mapping, or for tissue distribution studies. The  
XX CC present sequence represents a gene 216 cDNA sequence used in the scope of  
XX CC the invention  
XX SQ Sequence 207433 BP; 52775 A; 51289 C; 51698 G; 51671 T; 0 U; 0 Other;  
Query Match 2.5%; Score 35; DB 7; Length 207433;  
Best Local Similarity 55.3%; Pred. No. 66;  
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 1210 GGCCTAGCTAGATACATAGAGTCTATAGAAACATATAAGAGGCTTGAAGAACCAATTCAAAT 1269  
Db 1335 GTCTTACTCTTCTTCTTAAACACATACACAAATCCAGGGCTAGTAGGACCTTGAAT 1276  
QY 1270 CCTAAGGGTCTCTTCTTCTTGTGATCAGATCAAGAAATCATACACTCAAAACCGGAATC 1329  
Db 1275 ATACAGAACATCTCTTCTTAAAGAAACCTCAGGGATATAAAATCCAAATAGAAACCC 1216  
QY 1330 TTT 1332  
Db 1215 ATT 1213  
RESULT 12  
AAK84332

ID	XX	AAX84332 standard; DNA; 5059 BP.
AC	XX	AAX84332;
DT	XX	08-SEP-1999 (first entry)
DB	XX	Stealth virus nucleic acid clone, SEQ ID NO: 24.
KW	XX	Stealth virus; detection; diagnosis; infection; ss.
OS	XX	Stealth virus.
FH	FT	Key Location/Qualifiers
FT	FT	misc_difference 3605 /*tag= a
PT	PT	/note= "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	misc_difference 3610 /*tag= b
PT	PT	/note= "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	misc_difference 3615 /*tag= c
PT	PT	/note= "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	misc_difference 3630 /*tag= d
PT	PT	/note= "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	misc_difference 3631 /*tag= e
PT	PT	/note= "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	misc_difference 3638 /*tag= f
PT	PT	/note= "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	misc_difference 3641 /*tag= g
PT	PT	/note= "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	misc_difference 3642 /*tag= h
PT	PT	/note= "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	misc_difference 3657 /*tag= i
PT	PT	/note= "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	misc_difference 3659 /*tag= j
PT	PT	/note= "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	misc_difference 3681 /*tag= k
PT	PT	/note= "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	misc_difference 3698 /*tag= l
PT	PT	/note= "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"

```

FT misc_difference 3703
FT /tag= m
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3706
FT /tag= n
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3713
FT /tag= o
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3720
FT /tag= p
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3732
FT /tag= q
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3733
FT /tag= r
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3740
FT /tag= s
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3747
FT /tag= t
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3750
FT /tag= u
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3751
FT /tag= v
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3766
FT /tag= w
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3778
FT /tag= x
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3792
FT /tag= y
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3799
FT /tag= z
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"
FT misc_difference 3800
FT /tag= aa
FT /note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the base
FT numbering given in the specification"

```

[illegible]

D	b	3403	GYSARCHNGNGSNCSTHNDNATRCMCTRSSTATHVRSSSSSSSSSSSTHDSGNA	3466
Q	y	1147	CTTCATAGAGAGTGGCGCATTTGAAGTAGCTACCGCATCGAACCGGAAGCGGTTCAAGAC	1206
D	b	3463	TNRSTRTNSNCDBYTHNNGRGNHGWRSRSTTHNSHRTGNMSTTHSNCMRBTANDR	3522
Q	y	1207	ATGGGCGTACGTAGATACATAGAGTCATAGAACATAAAAGAGGCTTTGAAGAACCATTCA	1266
D	b	3523	MTHADSTRANHCWVANDARNTRGSNTNTHADSTRANETARBSNTNTHDSTRANHCWVHC	3582
Q	y	1267	AATCCTAAGGCTCTCTCTCTCTCTGTCATCACATCAAGAATCATACACTCAAACCAGGAA	1326
D	b	3583	VRNNSSTATHVRSCNSWTHNGFRNBTNBTNBTNBTNBTNBTNBTNBTNBTNBTNBTN	3642
Q	y	1327	CTCTTCTATCTTCCCTATAGCAATTC	1353
D	b	3643	BTBTCTCTCTCTCNBTATCBTBTB	3669
RESULT 13				
A	B	ABV74497	standard; DNA; 2799 BP.	
X	X	AC	ABV74497;	
X	X	AC	ABV74497;	
D	T	29-JAN-2003	(first entry)	
D	E	Human	type 1 diabetes receptive gene polymorphism 3 SEQ ID NO 7.	
K	W	Human;	type 1 diabetes receptive gene; gene; ds.	
O	S	Homo sapiens.		
P	N	JP2002238571-A.		
P	D	27-AUG-2002.		
P	F	15-FEB-2001;	2001JP-00038621.	
P	R	15-FEB-2001;	2001JP-00038621.	
P	A	{UYKY-}	UNIV KYUSHU.	
D	R	WPI;	2003-032789/03.	
P	T	Primers	for identifying of type 1 diabetes-receptive gene.	
P	S	Claim 7;	Page 9-10; 11pp; Japanese.	
C	C	The invention	relates to a primer pair used for identifying type 1	
C	C	diabetes	receptive gene. The present sequence is that of a polynucleotide	
C	C	expressing	a type 1 diabetes receptive gene polymorphism	
S	Q	Sequence	2799 BP; 601 A; 750 C; 880 G; 568 T; 0 U; 0 Other;	
Query Match 2.5%; Score 34.4; DB 7; Length 2799;				
Best Local Similarity 52.0%; Pred. No. 9.7;				
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;				
Q	y	875	TGCGGAAAACACCTTTAGATCTAGTTAAAGGGAAGCTGAAAGCTGAACCTGTCAGAAATA	934
D	b	81	TGGGGAGGACTGGGTGAAGGTAGAGAGAGGAGAACTTAACTTTAAGATGACAAATA	140
Q	y	935	AGCGCTGTGAATACAACTGTGTAIAACCCAAATTCAGTCGCTCAAGGGTGTCTGATATGCT	994
D	b	141	TTAAAGTAATTTTCAAGTACTCTTAAAAATGTAAATCTTTATCTATATATCTGATATGCA	200
Q	y	995	GGAGCTTCCCTGTCGCATTTGTGGGTAA	1022
D	b	201	AAATATTCATGACACTTGTGGAGTGA	228
RESULT 14				

		Query Match	2.5%	Score 34.4;	DB 8;	Length 32874;
		Best Local Similarity	52.0%;	Pred. No. 38;		
		Matches	77;	Conservative	0;	Mismatches 71; Indels 0; Gaps 0;
1237	AAACATAAAGGAGCTTGAAGAACGATTCAAACTCCTTAAGGGTCCTCTCTCTTCTTCTGCGATC	1296				
1173	AAACAACACAAAAATTTCTAAAAACCAATATAAACCCCAATRAAACTTACCAACTTTTAAAGCTC	1114				
1297	ACATCAAGAATCATACACTCAAAACGAGAACTCTTTCTCTATCTTCCCTCATAGCAATTCCTCA	1356				

QY

875

TCGCGAAAACACGTTTACATCTCTTAAAGGGAAGCTGAAGCTGACCTGTGAGAAATA

934

Db 8489 TGGGAGGACTGGGGTGNAGGTAGAGAGAGGAGGAGTTAACTTTAACTGACAAATA 8548  
Qy 935 AGCCTGTGGATACAAAGTTGATAACCCAAATTCAGTCGTCAGGGGTGCTGATATGCT 994  
Db 8549 TTTAAAGTATTTTCAAAGTACTCTTAAAAATGTAAATCTTTATCTATATATGATATGCA 8608  
Qy 995 GGACCTTCCCTGTCGCATTGCGGTAA 1022  
Db 8609 AATATTCATGACACCTTGGGATGA 8636

Search completed: June 10, 2004, 09:56:02  
Job time : 424.164 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 08:11:45 ; Search time 3792.71 Seconds  
(without alignments)  
15747.756 Million cell updates/sec

Title: US-10-070-386-1  
Perfect score: 1378  
Sequence: 1 ggcagctgggtgat.....accatcaatcaacctaaca 1378

Scoring table: IDENTIFY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rat.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1378	100.0	1378	6	BD013073	BD013073 Regulator
2	48.2	3.5	2000	6	AX655393	AX655393 Sequence
3	42.2	3.1	7218	6	I66494	I66494 Sequence 14
4	41.8	3.0	231585	2	AC103186	AC103186 Rattus no
5	41.8	3.0	329549	2	AC117331	AC117331 Rattus no
6	41.6	3.0	164377	2	AC136809	AC136809 Rattus no
7	41.6	3.0	226170	2	AC105687	AC105687 Rattus no
8	40	2.9	6197	6	AX344849	AX344849 Sequence
9	39.8	2.9	221649	2	AC137392	AC137392 Rattus no
10	39.8	2.9	262254	2	AC103016	AC103016 Rattus no
11	39.8	2.9	342084	2	AC128752	AC128752 Rattus no
12	39.6	2.9	81624	3	AC099763	AC099763 Caenorhab
13	39.6	2.9	290066	2	AC129421	AC129421 Rattus no
14	39.4	2.9	81921	10	AL805938	AL805938 Mouse DNA
15	39.4	2.9	186858	10	AC130210	AC130210 Mus muscu
16	39.4	2.9	195411	10	AL669958	AL669958 Mouse DNA
17	39.4	2.9	218807	2	BX511249	BX511249 Mus muscu
18	39	2.8	237119	10	AC115291	AC115291 Mus muscu
19	38.8	2.8	16512	3	CE049795	U49795 Caenorhab
20	38.8	2.8	34919	3	AF100663	AF100663 Caenorhab
21	38.8	2.8	192211	2	AC114644	AC114644 Mus muscu
22	38.6	2.8	129489	10	AL713860	AL713860 Mouse DNA
23	38.6	2.8	143148	10	AL607034	AL607034 Mouse DNA
24	38.6	2.8	233819	2	AC095098	AC095098 Rattus no
25	38.6	2.8	234873	2	AC137028	AC137028 Rattus no
26	38.6	2.8	240425	10	CNS07YOT	AL713839 Mus muscu
27	38.6	2.8	272563	2	AC095469	AC095469 Rattus no
28	38.4	2.8	195611	10	AC121971	AC121971 Mus muscu
29	38.4	2.8	334371	2	AC109571	AC109571 Rattus no
30	38.2	2.8	149092	2	AC127682	AC127682 Mus muscu
31	38.2	2.8	180907	9	AC131157	AC131157 Homo sapi
32	38.2	2.8	202275	2	AC102341	AC102341 Mus muscu
33	38.2	2.8	253768	2	AC125994	AC125994 Rattus no
34	38.2	2.8	302921	2	AC090998	AC090998 Homo sapi
35	38	2.8	75342	2	AC145944	Continuation (4 of
36	38	2.8	194370	2	AL645472	AL645472 Mus muscu
37	38	2.8	220638	10	AL645723	AL645723 Mouse DNA
38	38	2.8	239946	2	BX663615	BX663615 Danio rer
39	38	2.8	257760	2	AC126638	AC126638 Rattus no
40	37.8	2.7	114999	9	AF003421	AF003421 Homo sapi
41	37.8	2.7	142148	9	AC025465	AC025465 Homo sapi
42	37.8	2.7	167912	9	AC060232	AC060232 Homo sapi
43	37.8	2.7	170066	9	AL161730	AL161730 Human DNA
44	37.8	2.7	180001	9	AC112196	AC112196 Homo sapi
45	37.8	2.7	205224	10	AL671908	AL671908 Mouse DNA

ALIGNMENTS

RESULT 1	BD013073	1378 bp	DNA	linear	RAT 02-AUG-2002
LOCUS	BD013073	Regulatory sequences and expression system functional in mold			
DEFINITION	BD013073	Regulatory sequences and expression system functional in mold			
ACCESSION	BD013073.1	GI:22093262			
VERSION	BD013073.1	GI:22093262			
KEYWORDS	WO 0118219-A/1.				
SOURCE	unidentified				
ORGANISM	unidentified				
REFERENCE	1 (bases 1 to 1378)				
AUTHORS	Watanabe,M. and Murakami,T.				
TITLE	Regulatory sequences and expression system functional in mold fungi				
JOURNAL	Patent: WO 0118219-A 1 15-MAR-2001;				

Pred. No. is the number of results predicted by chance to have a



MEIJI SEIKA KAISHA LTD, MANABU WATANABE, TAKESHI MURAKAMI

**COMMENT**

OS	Myxelia sterilia	Key	Location/Qualifiers.
PN	W2 0118219 A/1		
PD	13-MAR-2001		
PF	07-SEP-2000 W0 2000JPD06104		
PF	07-SEP-1999 JP 99P 252B51		
PI	MANABU WATANABE, TAKESHI MURAKAMI		
PI	C12N15/80, C12N1/15, C12P21/00/((C12N15/80, C12R1:645), (C12N1/15,		
PC	C12R1:645),		
PC	(C12P21/00, C12R1:645)		
CC			
FH			

FH	Key	Location/Qualifiers.
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

```
Location/Qualifiers
1. .1378
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
```

## ORIGIN

```
Query Match      100.0%; Score 1378; DB 6; Length 1378;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	GTGAGCTGGGTGGTGATATCATGGCTGTGGTGTCCAAACTGTGTTAGTGTACAGAAAT	60
Db	1	GTGACGTGGGTGGTGATATCATGGCTGTGGTGTCCAAACTGTGTTAGTGTACAGAAAT	60
Qy	61	GAGGAAAGAAACGGTGGTGTGGCAGCTGAAGAAGAGAGGAGCCAAAGATAAATTC	120
Db	61	GAGGAAAGAAACGGTGGTGTGGCAGCTGAAGAAGAGAGGAGCCAAAGATAAATTC	120
Qy	121	ACAATGCCATACGGTTGGATCAATGCTGTTCGAAGAGAGCAGTTGGCATCTACCTGGTGT	180
Db	121	ACAATGCCATACGGTTGGATCAATGCTGTTCGAAGAGAGCAGTTGGCATCTACCTGGTGT	180
Qy	181	TCCCTCTTCCTGTGTACAAGATCAAGTATACGGATGACAGCCACCCCGCAACGGGAATCCG	240
Db	181	TCCCTCTTCCTGTGTACAAGATCAAGTATACGGATGACAGCCACCCCGCAACGGGAATCCG	240
Qy	241	GAGTTCAAAGAGGGTGTGCTACGGCAATTAGGTATAGATGGCATAGGGTTGACGTAA	300
Db	241	GAGTTCAAAGAGGGTGTGCTACGGCAATTAGGTATAGATGGCATAGGGTTGACGTAA	300
Qy	301	GCTGAAGAGCTGATTACGAGACATGAGACAACGAAAAATACAACGGTTGTATGGTTCCCGT	360
Db	301	GCTGAAGAGCTGATTACGAGACATGAGACAACGAAAAATACAACGGTTGTATGGTTCCCGT	360
Qy	361	GCTTACTAAAGTGATATCCAGAGACACACAGCGGAAGAAACCGATGCTCTGAGGGG	420
Db	361	GCTTACTAAAGTGATATCCAGAGACACACAGCGGAAGAAACCGATGCTCTGAGGGG	420
Qy	421	TTCCTTTAGAGTCTACATGTTAAAGTGCATGATAGAAACATCAAATGCCCAATCAAGTT	480
Db	421	TTCCTTTAGAGTCTACATGTTAAAGTGCATGATAGAAACATCAAATGCCCAATCAAGTT	480
Qy	481	AGTATACCTGACGCTACATCGCTTTCTTCGGGATCTTGCTTAAATATATGTGCTGTC	540
Db	481	AGTATACCTGACGCTACATCGCTTTCTTCGGGATCTTGCTTAAATATATGTGCTGTC	540
Qy	541	GAACTGTCGGTACTGCTTCGTACTAACTGTTCTTCCTGGTGAAGTCTTAGGACAAGCGCG	600
Db	541	GAACTGTCGGTACTGCTTCGTACTAACTGTTCTTCCTGGTGAAGTCTTAGGACAAGCGCG	600
Qy	601	CGTTTGTAGAGCTACATCATGTATGCCACATCTTAAAGCAGGATCTGAGACATTTCTAAGGC	660
Db	601	CGTTTGTAGAGCTACATCATGTATGCCACATCTTAAAGCAGGATCTGAGACATTTCTAAGGC	660
Qy	661	ATCCATATAGCGATTTGGCGCTAAGTCGCATTTGAAGAGATTAAGGGGGTGTGAAAGTG	720
Db	661	ATCCATATAGCGATTTGGCGCTAAGTCGCATTTGAAGAGATTAAGGGGGTGTGAAAGTG	720
Qy	721	GTGTGCTCAAAAGGAGGTCTGAATGGCTATACACCGCTTAAGCAGGTGGGCTAGCAGCTGT	780

Db	721	GTGTGTC	CAAAAGAGG	TGCGATT	TACCGACCGCT	TAAGCAGGTGGGCT	TAGCAGCTGT	780
Qy <td>781</td> <td>CTCAGCTG <td>GAATAAAC <td>GTGCAC <td>TCTTGGTATG <td>TCCACCTAAT <td>GTTCAGCAGATG <td>CAAA 840</td> </td></td></td></td></td></td>	781	CTCAGCTG <td>GAATAAAC <td>GTGCAC <td>TCTTGGTATG <td>TCCACCTAAT <td>GTTCAGCAGATG <td>CAAA 840</td> </td></td></td></td></td>	GAATAAAC <td>GTGCAC <td>TCTTGGTATG <td>TCCACCTAAT <td>GTTCAGCAGATG <td>CAAA 840</td> </td></td></td></td>	GTGCAC <td>TCTTGGTATG <td>TCCACCTAAT <td>GTTCAGCAGATG <td>CAAA 840</td> </td></td></td>	TCTTGGTATG <td>TCCACCTAAT <td>GTTCAGCAGATG <td>CAAA 840</td> </td></td>	TCCACCTAAT <td>GTTCAGCAGATG <td>CAAA 840</td> </td>	GTTCAGCAGATG <td>CAAA 840</td>	CAAA 840
Db <td>781</td> <td>CTCAGCTG <td>GAATAAAC <td>GTGCAC <td>TCTTGGTATG <td>TCCACCTAAT <td>GTTCAGCAGATG <td>CAAA 840</td> </td></td></td></td></td></td>	781	CTCAGCTG <td>GAATAAAC <td>GTGCAC <td>TCTTGGTATG <td>TCCACCTAAT <td>GTTCAGCAGATG <td>CAAA 840</td> </td></td></td></td></td>	GAATAAAC <td>GTGCAC <td>TCTTGGTATG <td>TCCACCTAAT <td>GTTCAGCAGATG <td>CAAA 840</td> </td></td></td></td>	GTGCAC <td>TCTTGGTATG <td>TCCACCTAAT <td>GTTCAGCAGATG <td>CAAA 840</td> </td></td></td>	TCTTGGTATG <td>TCCACCTAAT <td>GTTCAGCAGATG <td>CAAA 840</td> </td></td>	TCCACCTAAT <td>GTTCAGCAGATG <td>CAAA 840</td> </td>	GTTCAGCAGATG <td>CAAA 840</td>	CAAA 840
Qy <td>841</td> <td>TGCTGATT <td>TGGTTTAA <td>AAATGGC <td>CAATGTAG <td>TGTCGGGAA <td>CAAGCTTTAGATCT <td>AGTT 900</td> </td></td></td></td></td></td>	841	TGCTGATT <td>TGGTTTAA <td>AAATGGC <td>CAATGTAG <td>TGTCGGGAA <td>CAAGCTTTAGATCT <td>AGTT 900</td> </td></td></td></td></td>	TGGTTTAA <td>AAATGGC <td>CAATGTAG <td>TGTCGGGAA <td>CAAGCTTTAGATCT <td>AGTT 900</td> </td></td></td></td>	AAATGGC <td>CAATGTAG <td>TGTCGGGAA <td>CAAGCTTTAGATCT <td>AGTT 900</td> </td></td></td>	CAATGTAG <td>TGTCGGGAA <td>CAAGCTTTAGATCT <td>AGTT 900</td> </td></td>	TGTCGGGAA <td>CAAGCTTTAGATCT <td>AGTT 900</td> </td>	CAAGCTTTAGATCT <td>AGTT 900</td>	AGTT 900
Db <td>841</td> <td>TGCTGATT <td>TGGTTTAA <td>AAATGGC <td>CAATGTAG <td>TGTCGGGAA <td>CAAGCTTTAGATCT <td>AGTT 900</td> </td></td></td></td></td></td>	841	TGCTGATT <td>TGGTTTAA <td>AAATGGC <td>CAATGTAG <td>TGTCGGGAA <td>CAAGCTTTAGATCT <td>AGTT 900</td> </td></td></td></td></td>	TGGTTTAA <td>AAATGGC <td>CAATGTAG <td>TGTCGGGAA <td>CAAGCTTTAGATCT <td>AGTT 900</td> </td></td></td></td>	AAATGGC <td>CAATGTAG <td>TGTCGGGAA <td>CAAGCTTTAGATCT <td>AGTT 900</td> </td></td></td>	CAATGTAG <td>TGTCGGGAA <td>CAAGCTTTAGATCT <td>AGTT 900</td> </td></td>	TGTCGGGAA <td>CAAGCTTTAGATCT <td>AGTT 900</td> </td>	CAAGCTTTAGATCT <td>AGTT 900</td>	AGTT 900
Qy <td>901</td> <td>AAAGGGA <td>AGCTGAA <td>AGCTGAA <td>CTGTCAG <td>AAATAAG <td>CTGTTTGGAAAT <td>CAAACTGATAA 960</td> </td></td></td></td></td></td>	901	AAAGGGA <td>AGCTGAA <td>AGCTGAA <td>CTGTCAG <td>AAATAAG <td>CTGTTTGGAAAT <td>CAAACTGATAA 960</td> </td></td></td></td></td>	AGCTGAA <td>AGCTGAA <td>CTGTCAG <td>AAATAAG <td>CTGTTTGGAAAT <td>CAAACTGATAA 960</td> </td></td></td></td>	AGCTGAA <td>CTGTCAG <td>AAATAAG <td>CTGTTTGGAAAT <td>CAAACTGATAA 960</td> </td></td></td>	CTGTCAG <td>AAATAAG <td>CTGTTTGGAAAT <td>CAAACTGATAA 960</td> </td></td>	AAATAAG <td>CTGTTTGGAAAT <td>CAAACTGATAA 960</td> </td>	CTGTTTGGAAAT <td>CAAACTGATAA 960</td>	CAAACTGATAA 960
Db <td>901</td> <td>AAAGGGA <td>AGCTGAA <td>AGCTGAA <td>CTGTCAG <td>AAATAAG <td>CTGTTTGGAAAT <td>CAAACTGATAA 960</td> </td></td></td></td></td></td>	901	AAAGGGA <td>AGCTGAA <td>AGCTGAA <td>CTGTCAG <td>AAATAAG <td>CTGTTTGGAAAT <td>CAAACTGATAA 960</td> </td></td></td></td></td>	AGCTGAA <td>AGCTGAA <td>CTGTCAG <td>AAATAAG <td>CTGTTTGGAAAT <td>CAAACTGATAA 960</td> </td></td></td></td>	AGCTGAA <td>CTGTCAG <td>AAATAAG <td>CTGTTTGGAAAT <td>CAAACTGATAA 960</td> </td></td></td>	CTGTCAG <td>AAATAAG <td>CTGTTTGGAAAT <td>CAAACTGATAA 960</td> </td></td>	AAATAAG <td>CTGTTTGGAAAT <td>CAAACTGATAA 960</td> </td>	CTGTTTGGAAAT <td>CAAACTGATAA 960</td>	CAAACTGATAA 960
Qy <td>961</td> <td>CCCAATT <td>CACTGCT <td>CAAGG <td>GTCTCTG <td>ATATGCT <td>TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td> </td></td></td></td></td></td>	961	CCCAATT <td>CACTGCT <td>CAAGG <td>GTCTCTG <td>ATATGCT <td>TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td> </td></td></td></td></td>	CACTGCT <td>CAAGG <td>GTCTCTG <td>ATATGCT <td>TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td> </td></td></td></td>	CAAGG <td>GTCTCTG <td>ATATGCT <td>TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td> </td></td></td>	GTCTCTG <td>ATATGCT <td>TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td> </td></td>	ATATGCT <td>TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td> </td>	TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td>	TCGCATTTGGGGT 1020
Db <td>961</td> <td>CCCAATT <td>CACTGCT <td>CAAGG <td>GTCTCTG <td>ATATGCT <td>TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td> </td></td></td></td></td></td>	961	CCCAATT <td>CACTGCT <td>CAAGG <td>GTCTCTG <td>ATATGCT <td>TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td> </td></td></td></td></td>	CACTGCT <td>CAAGG <td>GTCTCTG <td>ATATGCT <td>TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td> </td></td></td></td>	CAAGG <td>GTCTCTG <td>ATATGCT <td>TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td> </td></td></td>	GTCTCTG <td>ATATGCT <td>TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td> </td></td>	ATATGCT <td>TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td> </td>	TGGAGCTTCCCTGT <td>TCGCATTTGGGGT 1020</td>	TCGCATTTGGGGT 1020
Qy <td>1021</td> <td>AACATTT <td>TCATAG <td>TGGGG <td>CAGAAAT <td>CAATCTAT <td>TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td> </td></td></td></td></td></td>	1021	AACATTT <td>TCATAG <td>TGGGG <td>CAGAAAT <td>CAATCTAT <td>TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td> </td></td></td></td></td>	TCATAG <td>TGGGG <td>CAGAAAT <td>CAATCTAT <td>TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td> </td></td></td></td>	TGGGG <td>CAGAAAT <td>CAATCTAT <td>TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td> </td></td></td>	CAGAAAT <td>CAATCTAT <td>TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td> </td></td>	CAATCTAT <td>TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td> </td>	TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td>	CTTAACTATTTCTG 1080
Db <td>1021</td> <td>AACATTT <td>TCATAG <td>TGGGG <td>CAGAAAT <td>CAATCTAT <td>TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td> </td></td></td></td></td></td>	1021	AACATTT <td>TCATAG <td>TGGGG <td>CAGAAAT <td>CAATCTAT <td>TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td> </td></td></td></td></td>	TCATAG <td>TGGGG <td>CAGAAAT <td>CAATCTAT <td>TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td> </td></td></td></td>	TGGGG <td>CAGAAAT <td>CAATCTAT <td>TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td> </td></td></td>	CAGAAAT <td>CAATCTAT <td>TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td> </td></td>	CAATCTAT <td>TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td> </td>	TTTCAATTGAA <td>CTTAACTATTTCTG 1080</td>	CTTAACTATTTCTG 1080
Qy <td>1081</td> <td>GTAGGAG <td>TTCTCA <td>ATGCTCT <td>TCGCTGT <td>CACATTA <td>CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td> </td></td></td></td></td></td>	1081	GTAGGAG <td>TTCTCA <td>ATGCTCT <td>TCGCTGT <td>CACATTA <td>CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td> </td></td></td></td></td>	TTCTCA <td>ATGCTCT <td>TCGCTGT <td>CACATTA <td>CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td> </td></td></td></td>	ATGCTCT <td>TCGCTGT <td>CACATTA <td>CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td> </td></td></td>	TCGCTGT <td>CACATTA <td>CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td> </td></td>	CACATTA <td>CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td> </td>	CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td>	TGCGGGGTCAACAACGT 1140
Db <td>1081</td> <td>GTAGGAG <td>TTCTCA <td>ATGCTCT <td>TCGCTGT <td>CACATTA <td>CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td> </td></td></td></td></td></td>	1081	GTAGGAG <td>TTCTCA <td>ATGCTCT <td>TCGCTGT <td>CACATTA <td>CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td> </td></td></td></td></td>	TTCTCA <td>ATGCTCT <td>TCGCTGT <td>CACATTA <td>CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td> </td></td></td></td>	ATGCTCT <td>TCGCTGT <td>CACATTA <td>CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td> </td></td></td>	TCGCTGT <td>CACATTA <td>CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td> </td></td>	CACATTA <td>CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td> </td>	CACACATCAT <td>TGCGGGGTCAACAACGT 1140</td>	TGCGGGGTCAACAACGT 1140
Qy <td>1141</td> <td>ATACAG <td>CTTCAT <td>AGAGTG <td>CGGCAT <td>TGAATG <td>AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td> </td></td></td></td></td></td>	1141	ATACAG <td>CTTCAT <td>AGAGTG <td>CGGCAT <td>TGAATG <td>AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td> </td></td></td></td></td>	CTTCAT <td>AGAGTG <td>CGGCAT <td>TGAATG <td>AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td> </td></td></td></td>	AGAGTG <td>CGGCAT <td>TGAATG <td>AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td> </td></td></td>	CGGCAT <td>TGAATG <td>AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td> </td></td>	TGAATG <td>AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td> </td>	AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td>	CGCATCGAACCCGGAAGCGTT 1200
Db <td>1141</td> <td>ATACAG <td>CTTCAT <td>AGAGTG <td>CGGCAT <td>TGAATG <td>AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td> </td></td></td></td></td></td>	1141	ATACAG <td>CTTCAT <td>AGAGTG <td>CGGCAT <td>TGAATG <td>AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td> </td></td></td></td></td>	CTTCAT <td>AGAGTG <td>CGGCAT <td>TGAATG <td>AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td> </td></td></td></td>	AGAGTG <td>CGGCAT <td>TGAATG <td>AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td> </td></td></td>	CGGCAT <td>TGAATG <td>AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td> </td></td>	TGAATG <td>AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td> </td>	AGCTAC <td>CGCATCGAACCCGGAAGCGTT 1200</td>	CGCATCGAACCCGGAAGCGTT 1200
Qy <td>1201</td> <td>CAAGAC <td>ATGCGG <td>CTGCT <td>AGATAC <td>ATAGAG <td>CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td> </td></td></td></td></td></td>	1201	CAAGAC <td>ATGCGG <td>CTGCT <td>AGATAC <td>ATAGAG <td>CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td> </td></td></td></td></td>	ATGCGG <td>CTGCT <td>AGATAC <td>ATAGAG <td>CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td> </td></td></td></td>	CTGCT <td>AGATAC <td>ATAGAG <td>CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td> </td></td></td>	AGATAC <td>ATAGAG <td>CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td> </td></td>	ATAGAG <td>CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td> </td>	CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td>	AAACATATAAAAGGAGCTTTGAAGAAC 1260
Db <td>1201</td> <td>CAAGAC <td>ATGCGG <td>CTGCT <td>AGATAC <td>ATAGAG <td>CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td> </td></td></td></td></td></td>	1201	CAAGAC <td>ATGCGG <td>CTGCT <td>AGATAC <td>ATAGAG <td>CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td> </td></td></td></td></td>	ATGCGG <td>CTGCT <td>AGATAC <td>ATAGAG <td>CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td> </td></td></td></td>	CTGCT <td>AGATAC <td>ATAGAG <td>CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td> </td></td></td>	AGATAC <td>ATAGAG <td>CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td> </td></td>	ATAGAG <td>CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td> </td>	CTCATAG <td>AAACATATAAAAGGAGCTTTGAAGAAC 1260</td>	AAACATATAAAAGGAGCTTTGAAGAAC 1260
Qy <td>1261</td> <td>CATTCA <td>AAATCCT <td>TAAGG <td>GTCTCT <td>CTTTCTG <td>ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td> </td></td></td></td></td></td>	1261	CATTCA <td>AAATCCT <td>TAAGG <td>GTCTCT <td>CTTTCTG <td>ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td> </td></td></td></td></td>	AAATCCT <td>TAAGG <td>GTCTCT <td>CTTTCTG <td>ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td> </td></td></td></td>	TAAGG <td>GTCTCT <td>CTTTCTG <td>ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td> </td></td></td>	GTCTCT <td>CTTTCTG <td>ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td> </td></td>	CTTTCTG <td>ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td> </td>	ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td>	CAAGATCATACACTCAAC 1320
Db <td>1261</td> <td>CATTCA <td>AAATCCT <td>TAAGG <td>GTCTCT <td>CTTTCTG <td>ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td> </td></td></td></td></td></td>	1261	CATTCA <td>AAATCCT <td>TAAGG <td>GTCTCT <td>CTTTCTG <td>ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td> </td></td></td></td></td>	AAATCCT <td>TAAGG <td>GTCTCT <td>CTTTCTG <td>ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td> </td></td></td></td>	TAAGG <td>GTCTCT <td>CTTTCTG <td>ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td> </td></td></td>	GTCTCT <td>CTTTCTG <td>ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td> </td></td>	CTTTCTG <td>ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td> </td>	ATCAGATCA <td>CAAGATCATACACTCAAC 1320</td>	CAAGATCATACACTCAAC 1320
Qy <td>1321</td> <td>CAGAACT <td>CTTTCT <td>TATCTT <td>CCCTAT <td>AGCAAT <td>TCCCAAA</td> <td>CCCATCAATCAACCTTAACA 1378</td> </td></td></td></td></td>	1321	CAGAACT <td>CTTTCT <td>TATCTT <td>CCCTAT <td>AGCAAT <td>TCCCAAA</td> <td>CCCATCAATCAACCTTAACA 1378</td> </td></td></td></td>	CTTTCT <td>TATCTT <td>CCCTAT <td>AGCAAT <td>TCCCAAA</td> <td>CCCATCAATCAACCTTAACA 1378</td> </td></td></td>	TATCTT <td>CCCTAT <td>AGCAAT <td>TCCCAAA</td> <td>CCCATCAATCAACCTTAACA 1378</td> </td></td>	CCCTAT <td>AGCAAT <td>TCCCAAA</td> <td>CCCATCAATCAACCTTAACA 1378</td> </td>	AGCAAT <td>TCCCAAA</td> <td>CCCATCAATCAACCTTAACA 1378</td>	TCCCAAA	CCCATCAATCAACCTTAACA 1378
Db <td>1321</td> <td>CAGAACT <td>CTTTCT <td>TATCTT <td>CCCTAT <td>AGCAAT <td>TCCCAAA</td> <td>CCCATCAATCAACCTTAACA 1378</td> </td></td></td></td></td>	1321	CAGAACT <td>CTTTCT <td>TATCTT <td>CCCTAT <td>AGCAAT <td>TCCCAAA</td> <td>CCCATCAATCAACCTTAACA 1378</td> </td></td></td></td>	CTTTCT <td>TATCTT <td>CCCTAT <td>AGCAAT <td>TCCCAAA</td> <td>CCCATCAATCAACCTTAACA 1378</td> </td></td></td>	TATCTT <td>CCCTAT <td>AGCAAT <td>TCCCAAA</td> <td>CCCATCAATCAACCTTAACA 1378</td> </td></td>	CCCTAT <td>AGCAAT <td>TCCCAAA</td> <td>CCCATCAATCAACCTTAACA 1378</td> </td>	AGCAAT <td>TCCCAAA</td> <td>CCCATCAATCAACCTTAACA 1378</td>	TCCCAAA	CCCATCAATCAACCTTAACA 1378
RESULT 2								
LOCUS	AX655393			2000 bp	DNA	linear		PAT 22-MAR-2003
DEFINITION	Sequence 5263 from Patent WO03000898.							
ACCESSION	AX655393							
VERSION	AX655393.1							
KEYWORDS	GI:29158207							
SOURCE	Oryza sativa							
ORGANISM	Oryza sativa							
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.							
REFERENCE	1							
TITLE	Chang, H. S., Chen, M., Cooper, B., Glazebrook, J., Goff, S. A., Hou, Y. M., Katagiri, P., Qian, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G. Plant genes involved in defense against pathogens							
JOURNAL	Patent: WO 03000898-A 5263 03-JAN-2003;							
FEATURES	Syngenta Participations AG (CH)							
source	Location/Qualifiers							
	1..2000							
	/organism="Oryza sativa"							
	/mol_type="unassigned DNA"							
	/db_xref="taxon:4530"							
ORIGIN								
Query Match	3.5%;							Score 48.2; DB 6; Length 2000;
Best Local Similarity	10.0%;							Pred. No. 0.0067;
Matches	79; Conservative 339; Mismatches 372; Indels 1; Gaps 1;							
ov	585	CCTAGG	CAAGG	CGGCTT	GTGTAG	ACCTAC	ATGATG	CCACATCTTTAAAGCAGGATCTG 644

```
Db 28 SCARMSSRRKMGSMKVKACSSCGKCKMTTRKSKWYSASSASGRTGSKWSSGYS 87
Qy 645 AGACATTTCTTAAGGCATCCATATAGGCATTTGGCGCTAAGTCGGCATTTGAAGGAGATA 704
Db 88 KGMKRRYKRSKRWGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGR 147
Qy 705 GGGGGGTGTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 764
Db 148 CYARGCGSCRRKSKGSGWGTCTCRGARGSGWSSGAKYKSGSMKRMWSSGCRSGCR 207
Qy 765 GTGGGCTAGCAGCTGTCTGAGCTGTGAATAAGCTCACCTTCCTTAGCTATGTCACCTAA 824
Db 208 RSAYRYGTGRKTYGTGKTYTYSASROMAYMTISWACSSYTWCRSKRSHWQWR 267
Qy 825 TGTGAGCAGATGCAAAATGCTGATTTGAGTTAAATGAGCATCTAGTGTAGTGTGCGAAGAC 884
Db 268 KMRWSRGYMSYMSYKMMCTAYKKSYYSCVYRGSGNRGATRYWGRGYMSRMWYK 327
Qy 885 ACCTTTAGATCTAGTTAAAGGAGCTGAAAGCTGAACCTGTGAGAAATAGCTGTGG 944
Db 328 KYWYRKYKMRKRWAGRMMSKRWKSKACYYTMRWMTTRRWAKKSRKSRK 387
Qy 945 AATACAAAGTTGATAACCAATTCAGTCGTCGCAAGGCTGTCTGATATGCTGGAGCTTCCC 1004
Db 388 RACMRKRYKMRGYSRMRSCKEARWKRCSRAWKMGCRGCMTCRMSKSYGMRKSW 447
Qy 1005 TGTGCGATTTGGGGTAATCTATTCATAGTGGGCGAGAAATGCAACTCTATTTCAATTGA 1064
Db 448 KRMASKYKMSRMYRMRKCKSRRTTMMGKTGGMGTGRCYKRGSGMRKCRKRWGR 507
Qy 1065 ATCTAAACTATTCGGGTAGGAGTCTCAATGGTCTTCTGCTGCTCACTTACACATCA 1124
Db 508 MYRWRKYRMYRBYTWYRCARKEYSYSAKARCACTRGKYWAGWMMRYKMYK 567
Qy 1125 TGGGGGTCAACAGCTATACAGCTTCA-TAGAGAGTCCGGCATTTGAAGTAGTACCCAT 1183
Db 568 MMYTKRYSKSWYCKMSYASCMKSAFKAAGAKCKESKMSANSKMSRSSRCKKCAKSR 627
Qy 1184 CGAACCCGAGCGGTTCAAGACATGGCGCTAGCATAGATAGATAGATAGATAGAT 1243
Db 628 SSARVYAMGWTSGSRMRKSVTCYWRKWSKSTCTWYMSKTYAKYSYWR 687
Qy 1244 AAAGGAGCTTGAAGAACCAATCAATCAAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1303
Db 688 RAMCWYWRWYTYRYSYMTYNAWYTSSTRAMTGMKYSGRYMTSWYKCKSKYRSMW 747
Qy 1304 GAATCATACACTCAACAGCAACTCTTTCTATCTTCCCTATAGCAATTCACAAACCCA 1363
Db 748 YYWSWAKTKMWRRYATRMWMMYRYSKMYTWCTMGMGYWYMTYMYKMYMYKCT 807
Qy 1364 TCAATCAACT 1374
Db 808 KTYWYNSATYW 818

RESULT 3
166494/c
LOCUS
DEFINITION
Sequence 14 from patent US 5670367.
ACCESSION
166494.1
VERSION
166494.1
KEYWORDS
Unknown.
SOURCE
Unknown.
ORGANISM
Unclassified.
REFERENCE
1 (bases 1 to 7218)
AUTHORS
Dorner, P., Scheifflinger, P. and Falkner, F. Gunter.
TITLE
Recombinant fowlpox virus
JOURNAL
Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
1..7218
/organism="unknown"
```

```
ORIGIN
Query Match 3.1%; Score 42.2; DB 6; Length 7218;
Best Local Similarity 8.9%; Pred. No. 0.48; Indels 0; Gaps 0;
Matches 41; Conservative 212; Mismatches 210;

Qy 607 TAGACCTACATGATGCCATCTTAAAGCAGGAGATCTGAGACATTTTCTAAGCATCAT 666
Db 1450 TAGAGAAATTGGTACTRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1391
Qy 667 ATAGCATTTGGCGCTAAGTCGGCATTTGAAGAGATAGGGGGGTGTGAAAGTGTGTGT 726
Db 1390 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1331
Qy 727 CAAAGGAGGTGCAATTTGGCTATACACGCGCTAAGCAGGTTGGGCTAGCAGCTGTCTG 786
Db 1330 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1271
Qy 787 CTGTGATAAAGCTCACCTTCTAGGTATGTCACCTAATGTGACGAGATGCAATGCTGA 846
Db 1270 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1211
Qy 847 TTGGGTTAAATGGGCGATGTAGTGTAGGTGCGCAAAACACAGTTAGATCTAGTTAAAGG 906
Db 1210 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1151
Qy 907 AAGCTGAAAGCTGCACTGTGAGAAATAGCCTGTTGGAATACAACTGTGATACCAAT 966
Db 1150 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1091
Qy 967 TCAGTCTCAAGCGGTCTGCTGATATCTGAGAGTTCCTCTGCTGCTGCTGCTGCTGCT 1026
Db 1090 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1031
Qy 1027 TCATAGTGGGCGAGAAATGCACTCTATTTCAATTTGAATCTA 1069
Db 1030 TTAATTTCTGAGCGTATGCGCAACGAGCAAAATAGTTATA 988

RESULT 4
AC103186
LOCUS
DEFINITION
Rattus norvegicus clone CH230-107D17, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.
AC103186
AC103186.5 GI:30578507
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
Mammalia; Eutheria; Rodentia; Chordata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa;
REFERENCE
1 (bases 1 to 231585)
AUTHORS
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benatmed, P.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Gantner, R., Garcia, A., Garner, T., Garza, M.,
Georgiadis, E., Geer, K., Gill, R., Grady, M., Guerr, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
```



## AUTHORS

Muzny, D. Marie., Metker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Centz, A., Chu, J., Chacko, J., Chavez, D., Chen, G., Chen, X., Chen, Z., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyie, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gueorgeogis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, J., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuwha, L., Loulseghe, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, D., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokeleneh, O., Okwunonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L., Puzos, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sittler, C. D., Smajda, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, R., Thomas, N., Thomas, D., Waldron, L., Walker, B., Wang, J., Valas, R., Vera, V., Villasana, D., Walgren, R., Wei, X., White, P., Wang, Q., Wang, S., Warren, J., Warren, R., Woodin, H., Worley, K., Williams, G., Willson, R., Wleczyk, R., Woodin, H., Worley, K., Wright, D., Wright, J., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiser, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

## TITLE

JOURNAL  
Unpublished  
2 (bases 1 to 329549)

## REFERENCE

Worley, K.C.  
Direct Submission

## AUTHORS

TITLE

JOURNAL

Submitted (10-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## REFERENCE

3 (bases 1 to 329549)

## AUTHORS

TITLE

JOURNAL

Submitted (09-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Sep 19, 2002 this sequence version replaced gi:21746236. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GUAC

Center clone name: CH230-242A22

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 198199 bases at least Q40

Consensus quality: 201002 bases at least Q30

Consensus quality: 202294 bases at least Q20

Estimated insert size: 210461; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 4 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 38119: contig of 38119 bp in length

\* 38120 38219: gap of unknown length

\* 38220 291497: contig of 253278 bp in length

\* 291498 291597: gap of unknown length

\* 291598 294124: contig of 2527 bp in length

\* 294125 294224: gap of unknown length

\* 294225 329549: contig of 35325 bp in length.

\* Location/Qualifiers

1. 329549

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-242A22"

complement(2196..2863)

/notes="clone\_boundary"

clone end:Sp6

site:MboI

and sequence:RXAAF11TV"

38220..39946

/note="wgs\_contig"

## ORIGIN

Query Match

Best Local Similarity 3.0%; Score 41.8; DB 2; Length 329549;

Mismatches 127; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 1110 CACTTACACATCATGGGGTCAACACGATATACAGCTTCATAGAGAGTGGCGCATTTGA 1169

DB 278081 CAGAGA 278140

QY 1170 AGTAGCTACCGCATCGAACCCGGAAGCGGTTCAAGACATCGGGCGGTACGTAGATACATAGA 1229

DB 278141 CACAGAGACACAGAGACACAGAGACACAGAGACACAGACACAGACACAGACACAGAGA 278200

QY 1230 GTCATAGAACATTAAGGAGCTTGNAGACCATCTCAATCTTAAGGGTCTCTCTCTTT 1289

DB 278201 GGCAGAGACACAGA 278260

QY 1290 CTGCATCATCAAGAAATCATACCTCAAAACAGGAACTCTTTCTATCTTCCCTATPAGCA 1349

DB 278261 CTGAGGAACATTTTAAAAAACATCATCACACAGTAAATTAACCAATCCAGAAATCTATA 278320

QY 1350 ATTCCCAAAACCCCATCAATCAACCTTACA 1378

Db 278321 CTTTCAATAAAGCTTCACTCACTTAA 278349

RESULT 6  
AC136809  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-349D12, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 8 unordered pieces.  
ACCESSION AC136809  
VERSION AC136809.2 GI:25188374  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 164377)  
AUTHORS Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Altschrocks, S., Amin, A., Augutano, D., Ayalew, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blakenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, J., Carter, K., Cavazos, I., Ceasar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, S., Davy-Carroli, L., De Anda, C., Dederich, D., Deigado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Li, J., Kwis, C., Kraft, C.L., Lebowitz, H., Levan, J., Lewis, J., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Louised, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, N., Mahindartne, M., Mahmoud, M., Mailloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M.P., McNeill, T.Z., Keenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nail, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwemet, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plapper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reich, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wlezyk, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE  
JOURNAL  
REFERENCE 2 (bases 1 to 164377)  
AUTHORS Rat Genome Sequencing Consortium.  
TITLE Direct Submission  
JOURNAL Submitted (08-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 164377)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (23-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 23, 2002 this sequence version replaced gi:24796682.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: KCSF  
Center clone name: CH230-349D12  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 119076 bases at least Q40  
Consensus quality: 123584 bases at least Q30  
Consensus quality: 126629 bases at least Q20  
Estimated insert size: 127842; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 3175: contig of 3175 bp in length  
\* 3176 3275: gap of unknown length  
\* 3276 115030: contig of 111755 bp in length  
\* 115031 115130: gap of unknown length  
\* 115131 119148: contig of 4018 bp in length  
\* 119149 119248: gap of unknown length  
\* 119249 155611: contig of 36363 bp in length  
\* 155612 155711: gap of unknown length  
\* 155712 159726: contig of 4015 bp in length  
\* 159727 159826: gap of unknown length  
\* 159827 161136: contig of 1310 bp in length  
\* 161137 161236: gap of unknown length  
\* 161237 162452: contig of 1216 bp in length  
\* 162453 162552: gap of unknown length  
\* 162453 164377: contig of 1825 bp in length.  
----- Location/Qualifiers  
1. 164377  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-349D12"  
----- misc\_feature  
1. 1297  
/note="wgs contig"  
----- misc\_feature  
1711. 3175  
/note="wgs contig"  
----- misc\_feature  
3276. 4359  
/note="wgs contig"

misc_feature	101466..103024								
	/notes="wgs_contig"								
	113307..115030								
	/notes="wgs_contig"								
	115131..116358								
misc_feature	/notes="wgs_contig"								
	119249..120860								
misc_feature	/notes="wgs_contig"								
	155712..157506								
misc_feature	/notes="wgs_contig"								
ORIGIN									
Query Match									
Best Local Similarity 56.8%; Pred. No. 0.81; Length 164377;									
Matches 77; Conservative 0; Mismatches 59; Indels 0; Gaps 0;									
QY	1110	CACATACACATCATCGGGGCTCAACAAAGTATACAGCTTCATAGAGAGTGGCGCATGGA	1169						
DB	103652	CACATACACATACATACAGGCACATTTAATTAACATACATACATACATGAGTGAGT	103711						
QY	1170	AGTAGTACCGCATCGAACCGGAGCGGTTCAGACATCGGGGTACGTAGTATACATACA	1229						
DB	103712	AGCATAGATCTCAATTCACACAAACAACTCATACAGACATACATACATACATACA	103771						
QY	1230	GTCTAGAAACATATAA	1245						
DB	103772	TACATACATACATACA	103787						
RESULT 7									
AC105687/C									
LOCUS									
DEFINITION Rattus norvegicus clone CH230-1111, *** SEQUENCING IN PROGRESS ***,									
3 unordered pieces.									
AC105687									
AC105687.3 GI:23101526									
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.									
KEYWORDS Rattus norvegicus (Norway rat)									
SOURCE Rattus norvegicus									
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;									
Rattus.									
1 (bases 1 to 226170)									
REFERENCE	Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,								
	Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,								
	Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,								
	Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,								
	Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,								
AUTHORS	Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,								
	Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,								
	Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,								
	Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,								
	Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,								
TITLE	Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,								
	Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,								
	Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,								
	Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,								
	Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,								
JOURNAL	Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,								
	Gunaratne, P., Haaland, K., Hamill, C., Hamilton, C., Hamilton, K.,								
	Harvey, V., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,								
	Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,								
	Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,								
COMMENT	Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,								
	Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, B., Kovar, C.,								
	Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,								
	Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,								
	Lorenshewa, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J.,								
JOURNAL	Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,								
	Mangum, B., Mapua, E., Martin, K., Martin, R., Martinez, E.,								
	Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, B.,								
	Milecavljevic, A., Miner, G., Minja, S., Montemayor, J., Moore, S.,								
	Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,								

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwackeme, O., Okwuonu, G., Olarpunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plapper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shateman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wiczyski, K., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.

Unpublished  
2 (bases 1 to 226170)  
Worley, K. C.  
Direct Submission  
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 226170)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Sep 18, 2002 this sequence version replaced gi:21743897.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GNCC  
Center clone name: CH230-1111  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 198186 bases at least Q40  
Consensus quality: 204314 bases at least Q30  
Consensus quality: 204704 bases at least Q20  
Estimated insert size: 229286; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved. 7584: contig of 7584 bp in length  
1



[illegible]

## COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center Project name: KZNG

Center clone name: CH230-unknown

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 150999 bases at least Q40

Consensus quality: 156151 bases at least Q30

Consensus quality: 159735 bases at least Q20

Estimated insert size: 155933; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 155377: contig of 155377 bp in length  
\* 155378 155477: gap of unknown length  
\* 155478 217856: contig of 62379 bp in length  
\* 217857 217956: gap of unknown length  
\* 217957 218983: contig of 1027 bp in length  
\* 218984 219083: gap of unknown length  
\* 219084 220305: contig of 1222 bp in length  
\* 220306 220405: gap of unknown length  
\* 220406 221649: contig of 1244 bp in length.

## FEATURES

source

1. 221649

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-unknown"

1. 22920

/notes="wgs contig"

32902..34790

/notes="wgs contig"

69822..72586

/notes="wgs contig"

89976..91515

/notes="wgs contig"

155478..156627

/notes="wgs contig"

172910..174635

/notes="wgs contig"

197144..200776

/notes="wgs contig"

## ORIGIN

Query Match 2.9%; Score 39.8; DB 2; Length 221649;  
Best Local Similarity 50.8%; Pred. No. 2.9;  
Matches 95; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 11 GGGTGATATCATGGCTGGTGTCCAAACACTGTTAGTACCAATGAGGAAGAA 70  
|||||  
Db 36230 GGTGTTGGATATCTCTCTTTGTTAGAAATATGTTAAAGAGAACATGAAAAAGGTCAGAG 36289  
QY 71 ACGGTGGTGTCTTGTGGCAGCTGAAGACTGAAGAGAGAGCCAAAGATATTCACAATCGCAT 130  
|||||  
Db 36290 AAGGCTTACAGAGGCTTCTACAAAAGGAAAAGCTTTCAACAAGAGAGAAAAGCTGA 36349  
QY 131 ACGTTGGATCATATGCTTGTTCAGAGAGACAGCTGCATCTACCTGCTGGTGTTCCTCTTC 190  
|||||  
Db 36350 AAGCTGGATCAAAAACCTGTTCTCCACCTCCCTGCTGTTGTCATCTGCTCCCTCTCT 36409  
QY 191 GTTGTAC 197  
|||||  
Db 36410 GTTGGAC 36416

## RESULT 10

AC103016

LOCUS

DEFINITION

Rattus norvegicus clone CH230-171119, \*\*\* SEQUENCING IN PROGRESS

\*\*\* 5 unordered pieces.

ACCESSION

AC103016

VERSION

AC103016.8 GI:30580739

KEYWORDS

HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus

REFERENCE

AUTHORS

Muzny,D.Marie, Metzker,M.Lee, Abramson,S., Adams,C., Alder,J.,

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,

Anyalebechi,V., Ayodeji,A., Ayodeji,M., Baca,E., Baden,H.,

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,

Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,

Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,

Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,

Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,

Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,M.,

Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

Gunaratne,P., Haaland,M., Hamill,C., Hamilton,C., Hamilton,K.,

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,

Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,

Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,D., Kovar,C.,

Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,

Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,

Lorensuewa,L., Loulaeeg,H., Lozada,R.J., Lu,X., Ma,J.,

Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,

Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,

Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,

Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,

Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,

Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,

Nwankweme,O., Okwuonu,G., Olarnpungoon,A., Pal,S., Parks,K.,

Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankuch,C.,

Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,

Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Resier,M.A., Reigh,R.,

Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,

Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,

Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,

Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,

Sneed,A., Sodergren,G., Song,X.-Z., Sorelle,R., Sosa,J.,

Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,

Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,

Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, C., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE  
JOURNAL

COURT  
REFERENCE  
AUTHORS  
2 (bases 1 to 262254)  
Unpublished  
Worley, K. C.

**AUTHORS**  
**TITLE**  
**JOURNAL**

JOURNAL  
SUBMITTED (24-NOV  
of Molecular and  
Baylor Plaza, Hou

## REFERENCE

AUTHORS	TITLE	Direct Submission	Rat Genome Sequencing Consortium.
JOHNSON	Submitted (12-MAY-2003)	Human Genome	

## JOURNAL

COMMENT

On May 13, 2003 this sequence version replaced G1:240123.1. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.basc.bcm.tmc.edu/>

web site: <http://www.hgsc.bcm.tmc.edu>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information

```

----- Project Information
Center project name: GIYH
Center clone name: CH230-271119

```

Center Name: CH230-E/ILIS  
----- Summary Statistics  
Assembly program: Atlas 3.0.

Assembly program: Atlas 3.0;  
Consensus quality: 213826 bases at least Q40  
Consensus quality: 217363 bases at least Q30

Consensus quality: 217363 bases at least Q30  
Consensus quality: 219200 bases at least Q20  
Estimated insert size: 225374; sum-of-con-

Estimated insert size: 225374; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

NOTE: This is a 'working draft' sequence. It currently (see [http://www.ngsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.ngsc.bcm.tmc.edu/docs/genbank_draft_data.html))

The true order of the pieces consists of 5 contigs. The order in this sequence record is not known and their order in this sequence record is different from the order in the contigs. Case between the contigs and the sequence record is different from the order in the contigs.

arbitrary. Gaps between the contigs are represented as  $\Delta$  but the exact sizes of the gaps are unknown.

\_\_\_\_\_

This record will be updated with the finished sequence as soon as it is available and the accession number will

1 251707: contig of 251707 bp in length

**251708** 251807: gap of unknown length  
**251808** 253055: contig of 1248 bp in length  
**252926**

253056 253155: gap of unknown length  
253156 254459: contig of 1304 bp in length  
254460 254459: contig of 1304 bp in length

254460 254559: gap of unknown length  
254560 256577: contig of 2018 bp in length  
254561 256577: contig of 2018 bp in length

256578 256677: gap of unknown length  
256678 262254: contig of 5577 bp in length.

Location/Qualifiers  
1. .262254

```
/organism="Rattus norvegicus"  
/mol_type="genomic DNA"
```

```

/db_xref="taxon:10116"
/clone="CH230-171519"
478. 1118
/note="clone_boundary
clone end:T7
site:EcoRI
end_sequence:BH288239"
complement(248949..24
/note="clone_boundary
clone end:Sp6
site:EcoRI
end_sequence:BH288241"

```

## ORIGIN

	Query Match	2.9%;	Score 39.8;	DB 2;	Length 262254;
	Best Local Similarity	50.8%;	Pred.No. 2.9;	92;	Indels 0; Gaps 0;
	Matches	95;	Conservative		
Qy	11	GTGGTGATATCATCGCTGTGGTGTCCAAAACCTGTGTAGTACTAGCAATGAGGAAGAA	70		
Db	31823	GTGTTTGGTACTTCTCTTTGTAGAAATATGGTTAAAGAGAAACATGAAGAGTCTGAG	3188		
Qy	71	ACGGTGGTGTGTGGCAGCTGAAGACTGAAGAGGAGCCAAAGATATTCACAAATCGAT	130		
Db	31883	AAGGCTTTAGACAAGGCTTTCTACAAAAGGAAAAGCTTTCAACAGAAAGAGAAAGAACTGA	3194		
Qy	131	ACGGTTGCATCAATGTGTTTCAAGAGAGACAGTGTGCATCTACTCTGGTTCCTCTTTC	190		
Db	31943	AAGCTGGAATCAAAACTGGTTTCTCCACCTCCCTTGGTTGTCCATCTGCTTCCCTCTCT	3200		
Qy	191	GTGTGAC	197		
Db	32003	GTGTGAC	32009		

RESULT 11

AC128752	342094 bp	DNA	linear	HTG 15-NOV-2002
LOCUS	Rattus norvegicus clone CH230-466L2,	WORKING DRAFT	SEQUENCE.	
DEFINITION	AC128752			
ACCESSION	AC128752.3	GI:25007641		
VERSION	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.			
KEYWORDS	Rattus norvegicus (Norway rat)			
SOURCE	Rattus norvegicus			
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
REFERENCE	1 (bases 1 to 342084)			
AUTHORS	Muzny,D,Marie, Metzker,M, Lee, Abramson,S., Adams,C, Alder,J.,			

Buzay, D. Marie, Retzke, S. Bee, Adamson, J. S., Reamer, S. J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angulano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Banderanaika, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, E., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Durn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregiorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kratt, C. L., Labow, H., Levan, J., Lewis, L. I. Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzshewa, J., Louisegh, H., Lozado, R. J., Lu, X., Ma, J.,

Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwako, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Picpper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steelme, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmari, K., Valas, R., Vera, V., Villanueva, D., Waldron, D., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhauser, A., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

Direct Submission  
 Unpublished  
 2 (bases 1 to 342084)  
 Worley, K.C.

Direct Submission  
 Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 342084)  
 Rat Genome Sequencing Consortium.

Direct Submission  
 Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Nov 15, 2002 this sequence version replaced gi:23803473. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: KALG  
 Center clone name: CH230-46612  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 143887 bases at least Q40  
 Consensus quality: 144547 bases at least Q30  
 Consensus quality: 144951 bases at least Q20  
 Estimated insert size: 177730; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes

\* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 342084: contig of 342084 bp in length.

FEATURES  
 Location/Qualifiers  
 source  
 1..342084  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-46612"  
 misc\_feature  
 184048..186150  
 /note="wgs\_contig"  
 ORIGIN

Query Match 2.9%; Score 39.8; DB 2; Length 342084;  
 Best Local Similarity 50.8%; Pred.No. 3;  
 Matches 95; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 11 GTGGTATATCATGGCTGTGTCCTTCTTTAGATATGTTAAAGAGAACATGAAAAGGTCAGAG 90861  
 Db 90802 GTGTTGGTATATCTCTCTTTCTTTAGATATGTTAAAGAGAACATGAAAAGGTCAGAG 90861  
 QY 71 ACGGTGGTATCATGCTTCTTCAAGAGACGCTTGCATCTACCTGGTGTTCCTCTCTTC 130  
 Db 90862 AAGGCTTAGACAGGCTTCTACAAAAGGAAAGCTTTCAACAGAAAGAAAGCTGA 90921  
 QY 131 ACGGTGGTATCATGCTTCTTCAAGAGACGCTTGCATCTACCTGGTGTTCCTCTCTTC 190  
 Db 90922 AAGCTGGAATCAAACTGTTCTCCACCTCCCTTGGTGTTCCTCTCTCTCTCTCTCT 90981  
 QY 191 GTTGATAC 197  
 Db 90982 GTTGATAC 90988

RESULT 12  
 AC099763/c  
 LOCUS  
 DEFINITION  
 AC099763  
 AC099763.1 GI:17017635  
 VERSION  
 HTG.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Caenorhabditis briggsae  
 Caenorhabditis briggsae  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 81624)  
 Washington University Genome Sequencing Center.  
 The C. briggsae Genome Sequencing Project  
 Unpublished  
 2 (bases 1 to 81624)  
 Waterston, R.  
 Direct Submission  
 Unpublished  
 3 (bases 1 to 81624)  
 Waterston, R.  
 Direct Submission  
 Submitted (20-NOV-2001) Department of Genetics, Washington University, Genome Sequencing Center, 444 Forest Park Avenue, St. Louis, MO 63110, USA  
 Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University  
 St. Louis, MO 63110, USA  
 email: [rw@nematoe.wustl.edu](mailto:rw@nematoe.wustl.edu)

COMMENT  
 NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

#### FEATURES

source

Location/Qualifiers  
1..81624  
/organism="Caenorhabditis briggsae"  
/mol\_type="genomic DNA"  
/strain="Gujarat G16"  
/db\_xref="taxon:6238"

#### ORIGIN

Query Match 2.9%; Score 39.6; DB 3; Length 81624;  
Best Local Similarity 58.5%; Pred No. 3.2;  
Matches 69; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
  
QY 34 TCANAACTGTTAGTACGATGAGAAAGAACGGTGGTGTGGCAGCTGAA 93  
DB 8467 TTTCAAAATATCTATTATTATGCGAGCTGTAAAAAAGAAAGTTTGTACCAATTGAG 8408  
  
QY 94 GACTGAGAGGAGCCAGACATATTCACATCGATACGGTTGTCATCAATGCTTCT 151  
DB 8407 ACATGAAATACGTATTAATTAATATCAGATGAGCAAGATGATGATGTTAGTT 8350

RESULT 13  
AC129421/c  
LOCUS  
DEFINITION  
AC129421 290066 bp DNA linear HTG 20-NOV-2002  
Rattus norvegicus clone CH230-62C13, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 5 unordered pieces.  
AC129421  
VERSION  
AC129421.3 GI:25138103  
KEYWORDS  
HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
SOURCE  
Rattus norvegicus (Norway rat)  
ORGANISM  
Rattus norvegicus  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

#### REFERENCE

1 (bases 1 to 290066)  
Muzny, D., Marie, H., Metzger, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gregorogis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hu, Y.K., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loresukewa, L., Loulsged, H., Lozado, R., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, C., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okunolu, G., Oiarupunsagoon, A., Pal, S., Parks, K.,

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats, all regions were covered by at least one plasmid subclone or more than one M13 subclone, and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-248L9 is from the RPI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACel.6.  
Location/Qualifiers  
1. 81921  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/clone="RP23-248L9"  
/clone\_lib="RPI-23"

FEATURES

source  
1. 81921  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/clone="RP23-248L9"  
/clone\_lib="RPI-23"

ORIGIN

Query Match 2.9%; Score 39.4; DB 10; Length 81921;  
Best Local Similarity 51.4%; Pred. No. 3.7;  
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 546 GTGGTACTGCTTCTACTACTGTTCTTCGTTCAAGTCTTAGCACAAGCCCGGTTT 605  
DB 81899 GTTGACACACTTTGTGCAACAGCCCTCTACTATCTTCTTAGGAGATATCTGATT 81830  
QY 606 GTAGACCTACATGTCACATCTTTAAAGCGGGATCTGAGACATTTCTTAAGGATCCA 665  
DB 81829 GGGTACCTAAATGCTGCAGAACAGTAATAATATTCACCTTTTCTCTCTGACATACA 81770  
QY 566 TATAGGCTTGGCGCTAAGTCGCAATTAAGAGAGATGAAGGGGTGTGAAAGTGT 722  
DB 81769 TAACAGCACTGATTCAGCAGTCCGAATGTGTGTAGCAGAGAAAGACGAGTGT 81713

RESULT 15

AC130210 186858 bp DNA linear ROD 27-NOV-2003

AC130210 Mus musculus BAC clone RP24-344N22 from chromosome 7, complete sequence.

AC130210 AC130210.4 GI:30231364

HTG. Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Nguyen, C. and Kozlowicz, A.

The sequence of Mus musculus BAC clone RP24-344N22

Unpublished (2001)

2 (bases 1 to 186858)

Wilson, R.

Sequencing of Mus musculus

Unpublished (2001)

3 (bases 1 to 186858)

McPherson, J.D. and Waterston, R.H.

Direct Submission

Submitted (03-APR-2003) Genome Sequencing Center, 4444 Forest Park

5 (bases 1 to 186858)

McPherson, J.D. and Waterston, R.H.

LOCUS

AC130210/c

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

RESULT 14

AL805938/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL805938 81921 bp DNA linear ROD 26-SEP-2002

Mouse DNA sequence from clone RP23-248L9 on chromosome X, complete sequence.

AL805938 AL805938.4 GI:23320936

HTG. Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Tromans, A.

Direct Submission

Submitted (24-SEP-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [humquerry@sanger.ac.uk](mailto:humquerry@sanger.ac.uk) Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)

On Sep 25, 2002 this sequence version replaced gi:22265512.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humquerry@sanger.ac.uk](mailto:humquerry@sanger.ac.uk)

-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate

\* 281951 282050: gap of unknown length  
\* 283370: contig of 1320 bp in length  
\* 283371 283470: gap of unknown length  
\* 283471 286076: contig of 2606 bp in length  
\* 286077 286176: gap of unknown length  
\* 286177 290066: contig of 3890 bp in length.

FEATURES

source

1. 290066  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-62C13"  
32547..33462  
/note="clone boundary  
clone\_end:Sp6  
site:  
end sequence: BH287162"  
40710..42036  
/note="wgs\_contig"  
280810..281950  
/note="wgs\_contig"

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature



```
TITLE      Direct Submission
JOURNAL    Submitted (30-APR-2003) Genome Sequencing Center, 4444 Forest Park
REFERENCE  Parkway, St. Louis, MO 63108, USA
AUTHORS    6 (bases 1 to 186858)
TITLE      Wilson,R.
JOURNAL    Direct Submission
COMMENT    Submitted (27-NOV-2003) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           On Apr 30, 2003 this sequence version replaced gi:29501927.
           ----- Genome Center
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu
           Contact: submissions@wustl.edu
           ----- Summary Statistics
           Center project name: M_BB0344N22
           -----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC library has been constructed by Pieter de Jong and
coworkers (http://www.Chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.Chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1. .186858
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="7"
  /map="7"
  /clone="RP24-344N22"
  /clone_lib="RPCI-24"
  1334. .1484
    /rpt_family="Alu"
  4834. .4946
    /rpt_family="Alu"
  5035. .5180
    /rpt_family="B2"
  6535. .6651
    /rpt_family="L1"
  7259. .7406
    /rpt_family="B2"
  7432. .7567
    /rpt_family="Alu"
  7598. .7663
    /rpt_family="B2"
  7664. .7780
    /rpt_family="B4"
  7845. .8055
    /rpt_family="B2"
  8235. .8338
    /rpt_family="B2"

repeat_region 9419. .9580 /rpt_family="MER2_type"
repeat_region 10335. .10537 /rpt_family="B4"
repeat_region 11210. .11361 /rpt_family="B2"
repeat_region 11411. .11497 /rpt_family="B4"
repeat_region 11810. .11962 /rpt_family="Alu"
repeat_region 12046. .12433 /rpt_family="B4"
repeat_region 12680. .12727 /rpt_family="ERVK"
repeat_region 12774. .12972 /rpt_family="ERV1"
repeat_region 13385. .13502 /rpt_family="B2"
repeat_region 13657. .13857 /rpt_family="Alu"
repeat_region 14075. .14147 /rpt_family="B2"
repeat_region 14201. .14226 /rpt_family="ID"
repeat_region 14227. .14357 /rpt_family="B4"
repeat_region 14479. .14860 /rpt_family="Alu"
repeat_region 17823. .17954 /rpt_family="L1"
repeat_region 18495. .18597 /rpt_family="Alu"
repeat_region 18545. .18631 /rpt_family="Alu"
repeat_region 18973. .19078 /rpt_family="B2"
repeat_region 19119. .19242 /rpt_family="B4"
repeat_region 19243. .19352 /rpt_family="Alu"
repeat_region 19432. .19499 /rpt_family="B4"
repeat_region 22725. .22836 /rpt_family="ID"
repeat_region 23994. .24053 /rpt_family="B4"
repeat_region 24034. .24194 /rpt_family="B4"
repeat_region 24230. .24487 /rpt_family="B2"
repeat_region 26457. .26650 /rpt_family="L1"
repeat_region 28112. .28276 /rpt_family="B2"
repeat_region 29623. .29699 /rpt_family="MER1_type"
repeat_region 29973. .30306 /rpt_family="L2"
repeat_region 30609. .30705 /rpt_family="ERVK"
repeat_region 31040. .31148 /rpt_family="Alu"
repeat_region 31927. .31987 /rpt_family="B4"
repeat_region 32001. .32082 /rpt_family="Alu"
repeat_region 32158. .32257 /rpt_family="L1"
repeat_region 32837. .32963 /rpt_family="L1"
repeat_region 33965. .34087 /rpt_family="Alu"
repeat_region 33965. .34087 /rpt_family="B2"
```



Search completed: June 10, 2004, 11:49:37  
Job time : 3799.21 secs

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 09:32:10 ; Search time 2139.6 Seconds

(without alignments)  
15199.092 Million cell updates/sec

Title: US-10-070-386-2

Perfect score: 1089

Sequence: 1 taaactccatcatatagcgg.....actgcaatgagtcgcccgg 1089

Scoring table: IDENTITY\_NUC

Gapop 10\_0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rtd:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42.8	3.9	1101	29	CNS0000D1
C	2	42.2	3.9	1201	13 BX376097
C	3	41	3.8	440	28 A0634841
C	4	40.2	3.7	1101	29 CNS0039G

C	5	40	3.7	1101	29	CNS014QD
		39.8	3.7	638	28	AQ385240
		39	3.6	807	28	BZ718506
		39	3.6	833	28	B2417799
		39	3.6	894	29	CG107311
		39	3.6	973	29	CG038009
		39	3.6	505	14	CE943651
		38.8	3.6	505	14	CE943651
		38.4	3.5	492	29	CE323888
		38.2	3.5	823	13	BU374247
		38.2	3.5	839	29	CG865000
		37.8	3.5	586	10	AW384856
		37.8	3.5	595	13	BW240301
		37.6	3.5	1201	13	BX337130
		37.4	3.4	555	28	AZ943190
		37.4	3.4	660	9	A1488706
		37.4	3.4	743	9	A1486681
		37.4	3.4	842	29	ECAS76716
		37.2	3.4	581	29	CE457309
		37.2	3.4	760	29	CG120814
		37.2	3.4	797	29	CNS00358
		37	3.4	458	29	EX002272
		37	3.4	552	13	BU497199
		37	3.4	624	29	CE610094
		37	3.4	1001	29	CNS01400
		36.8	3.4	524	14	CF180840
		36.8	3.4	560	28	AQ237122
		36.8	3.4	657	28	CC095789
		36.8	3.4	762	29	AG030225
		36.6	3.4	353	10	BF490341
		36.6	3.4	355	10	BE978369
		36.6	3.4	368	28	BZ761149
		36.6	3.4	413	14	CB207284
		36.6	3.4	587	10	AW441466
		36.4	3.3	366	10	AW765028
		36.4	3.3	411	12	BG553280
		36.4	3.3	442	10	BG023598
		36.4	3.3	481	12	BG413804
		36.4	3.3	490	12	BG348032
		36.4	3.3	492	13	BQ737566
		36.4	3.3	531	12	BG579410
		36.4	3.3	612	14	CF287145

#### ALIGNMENTS

RESULT 1	CNS0000D1	1101 bp	DNA	linear	GSS 03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence TET3 end of BAC #				
DEFINITION	BACK01J16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL065414				
VERSION	AL065414.1	GI:4938827			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster (fruit fly)				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseogawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of				



REFERENCE 1 (bases 1 to 440)  
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.  
 TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building  
 JOURNAL Map Building (1997)  
 COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbeetigr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source  
 1..440  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="GBA:7683074"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-477M3"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /notes="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI; RPCI11 Human Male BAC Library"

ORIGIN  
 Query Match 3.8%; Score 41; DB 28; Length 440;  
 Best Local Similarity 50.8%; Pred. No. 0.46;  
 Matches 95; Conservative 0; Mismatches 92; Indels 0; Gaps 0;  
 QY 640 CACTGCTATGCTGGGTATGTAATCTCGCTAGTTCATTAAGGGGACATAGAAACAATA 699  
 DB 401 CATTGTGTCTTAAGCCACNAGCTCTGTGTATTTTGTATGGAGCTTAAGAAACTA 342  
 QY 700 CTTCACTGTCAGATCAGTAAGCTCTGTACATTTCTCACTGCAAAATTTCTGAACAATGGAG 759  
 DB 341 ACACAACTACAGAAAGGTAGTCTGTATCGCTTACAGAACATCACTATAAAATAGAAC 282  
 QY 760 AGCATTATGAATTAATGGAATCTCTCATTTATAGTGAAGGAAACAGAGCGCCCTTTA 819  
 DB 281 AACTGCTANAAGATCTTTTGGTCTTTTATCAAAAGTGAGCAACAATGAAGACTTATA 222  
 QY 820 TTATGAA 826  
 DB 221 TTATCAA 215

RESULT 4  
 CNS0039G/c 1101 bp DNA linear GSS 03-JUN-1999  
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
 DEFINITION BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
 ACCESSION AL063921.1 GI:4941778  
 VERSION AL063921.1  
 KEYWORDS GSS.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 1101)  
 Genoscope.

REFERENCE 1 (bases 1 to 440)  
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.  
 TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building  
 JOURNAL Map Building (1997)  
 COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbeetigr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source  
 1..1101  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone="BACR08K10"  
 /clone\_lib="RPCI-98"  
 /note="end : TET3"

ORIGIN  
 Query Match 3.7%; Score 40.2; DB 29; Length 1101;  
 Best Local Similarity 14.7%; Pred. No. 1.4;  
 Matches 60; Conservative 184; Mismatches 163; Indels 0; Gaps 0;  
 QY 327 TGAATGGGTATTCATATATATGAGCTTGCACCCGCCCAATATGTAAACAGACCAACAATG 386  
 DB 940 WDDWDWDKWDGAKDKKADDDDDGAGDKDDGKDDADDGTDGTDKDDDKDDWDKA 881  
 QY 387 GGTCACTGGGCCCCCAGACATGTGGCTATATAAGTACCTCTGTAGCAATCAGACTTAC 446  
 DB 880 KGTWGDATWAAATDWWGWDADWWTADAAADDDWADDDRDWDAWAWKWDADAWGART 821  
 QY 447 TGATAGACGCTCCCTATATATGCTATAAATAAGTCACCTACTAGAACTACCGACAGTGTG 506  
 DB 820 DRDWDGAGKGGGARKRRDRKRDADDDADDDADATTTTATTTTDDDDKWKTD 761  
 QY 507 AAATCCGACAGTGTCTGTGTGAACATCTCATGTCTATATGAATGAATGAAGAAG 566  
 DB 760 WTRWAADRTWRDDDDDDRDAGTAGCKWRRTWKRRKRRDTRWDDADADDTARDRRRG 701  
 QY 567 GTGTACGGGTAGTACGAATCTGTATGATATCAATGAGTACGCTAGTGTGAACAGCG 626  
 DB 700 DDGADAGKGTGKRRRRDRATWDRDADAAWTTTDDDDWDRRRRRKARR 641  
 QY 627 GATCGGATCTAGCACTGCTATGCTGGGTATGTAATCTCTGCTATGTTCTAAGGCGA 686  
 DB 640 RETTARAADWWTKAWDKNWDMKTRADRDWDAADTTWTDARKADRDWAKARARRD 581  
 QY 687 CATGAAGAATACCTCAGTGTGACATACGTAAGCTCTGTACATT 733  
 DB 580 RARAARDRRTTKGTTTATTTWTAARAAMWMAWMAWATTATTTT 534

RESULT 5  
 CNS014QD/c 1101 bp DNA linear GSS 26-JUL-1999  
 LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC  
 DEFINITION BACN12C15 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
 ACCESSION AL104479  
 VERSION AL104479.1 GI:5616202  
 KEYWORDS GSS.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster



```

Seq primer: TP
Class: sheared ends.
Location/Qualifiers
  1. .807
    /organism="Zea mays"
    /mol_type="genomic DNA"
    /strain="B73"
    /db_xref="taxon:4577"
    /clone="ZMMBTal36L20"
    /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high"
    Cor selected genomic DNA library"

ORIGIN
Query Match          3.6%; Score 39; DB 28; Length 807;
Best Local Similarity 48.8%; Pred. No. 2.7;
Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 579 AGTACGAATCTGTATGATATCAATGTCAGTGTGTAACAGCGGATCGGATCTA 638
    |||||
Db 115 AGTACGAATGTTACATATATATTTAAAGTCTGTAAATCAAGGGTACTTTGAA 174

QY 639 GCATCTGCTATGCTGGGTATGTAATCTCGCTATGTTCTATPAGGGGCGACATAGAAAGAT 698
    |||||
Db 175 GCACCAATATTTTGGAAACCAATCAAAACATCATTTAAATTTAGACAAACAGAGGTAT 234

QY 699 ACCTCAGTGTGACATACGTATGCTGTACATTTCACTGCAATTTCTGAAATTTGA 758
    |||||
Db 235 AACTCAGTGTGCAATGCTCTTTTAAAGCTCATCAGCACCTGAACATTAAGA 294

QY 759 GAGCATTATGAATACTAAATGGAATCTCTCATTA 793
    |||||
Db 295 AAGAGGTTCACAGTCAGCAATACAGACTTCAGTA 329

RESULT 8
BZ417799
LOCUS          833 bp      DNA      linear      GSS 10-DEC-2002
DEFINITION    if81f07-g2 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
KEYWORDS      if81f07 5', genomic survey sequence.
ACCESSION     BZ417799
VERSION       BZ417799.1 GI:26361596
SOURCE        Zea mays
ORGANISM      Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 833)
Rabinowicz,P.D., O'Shaughnessy,A.L., Ballja,V., Dedhia,N.,
Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
Zutavern,T., McCombie,W.R. and Martienssen,R.A.
Genomic shotgun sequences from Zea mays (methyl-filtered)
Unpublished (2002)
Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org
Plate: if81 row: f column: 07
Seq primer: -21M13Univrev
Class: shotgun
High quality sequence stop: 833.
Location/Qualifiers
  1. .833
    /organism="Zea mays"
    /mol_type="genomic DNA"
    /cultivar="B73"
    /db_xref="taxon:4577"
    /clone="if81f07"
    /lab_host="DH5a"

FEATURES
  source
    1. .833
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="if81f07"
      /lab_host="DH5a"

/clone lib="WGS-ZmaysF (DH5a methyl filtered)"
/note="Organ: immature ears; Site 1: Xba 1; Site 2: Xba 1;
The vector was digested with XbaI and one nucleotide was
added by fill in the recessive 3' end. The genomic DNA
was nebulized, end repaired, adaptor ligated and size
fractionated using sephadex. The resulting fragments were
between 0.8 and 3 kb and were cloned into the vector (.x/y
reads in M13mp19, .b/g reads in pUC19). The same ligation
was transformed into DH5a."

ORIGIN
Query Match          3.6%; Score 39; DB 28; Length 833;
Best Local Similarity 48.8%; Pred. No. 2.8;
Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 579 AGTACGAATCTGTATGATATCAATGTCAGTGTGTAACAGCGGATCGGATCTA 638
    |||||
Db 604 AGTACGAATGTTACATATATATTTAAAGTCTGTAAATCAAGGGTACTTTGAA 663

QY 639 GCATCTGCTATGCTGGGTATGTAATCTCGCTATGTTCTATPAGGGGCGACATAGAAAGAT 698
    |||||
Db 664 GCACCAATATTTTGGAAACCAATCAAAACATCATTTAAATTTAGACAAACAGAGGTAT 723

QY 699 ACCTCAGTGTGACATACGTATGCTGTACATTTCACTGCAATTTCTGAAATTTGA 758
    |||||
Db 724 AACTCAGTGTGCAATGCTCTTTTAAAGCTCATCAGCACCTGAACATTAAGA 783

QY 759 GAGCATTATGAATACTAAATGGAATCTCTCATTA 793
    |||||
Db 784 AAGAGGTTCACAGTCAGCAATACAGACTTCAGTA 818

RESULT 9
CG107311/c
LOCUS          834 bp      DNA      linear      GSS 20-AUG-2003
DEFINITION    PUJEA72TB ZM 0.6-1.0 KB zea mays genomic clone ZMMETA0647K23,
Genomic survey sequence.
ACCESSION     CG107311
VERSION       CG107311.1 GI:33990189
KEYWORDS      GSS.
SOURCE        Zea mays
ORGANISM      Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 894)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PUJEA72TD
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
Location/Qualifiers
  1. .894
    /organism="Zea mays"
    /mol_type="genomic DNA"
    /strain="B73"
    /db_xref="taxon:4577"
    /clone="ZMMETA0647K23"
    /clone lib="ZM 0.6-1.0 KB"
    /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high"
    Cor selected genomic DNA library"

FEATURES
  source
    1. .894
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMETA0647K23"
      /clone lib="ZM 0.6-1.0 KB"
      /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high"
      Cor selected genomic DNA library"

```

Best Local Similarity 48.8%; Pred. No. 2.9;  
Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 579 AGTACAACTGTATGATATCAATCGTAGCAGTGTGTAACACGCGGATCGGATCTA 638  
Db 225 AGTAACAAATGATGATATATATATTTTAAGTCTGTAATCAAGGCTACTTTTGA 166  
QY 639 GCACCTGATGTCTGGTATGTAATCTGCTATGTCATAGCGGCACATAGAAAGAT 698  
Db 165 GCACCATATTTTGGAACTAATTCACATCATTTAATTAGACAAACAGAGCTAT 106  
QY 699 ACTGAGTGCACATGATGATCTGTACATTTCACTGCGAAATTTCTGAAACATGGA 758  
Db 105 AACTCACTGTCGAACTGCAATGCTCTCTTTTAAAGTCTCATCGCACCTGAAACATTAAGA 46  
QY 759 GACCATTTGMAATCAATAATGAAGTCTCTCATTA 793  
Db 45 AAGAGTCCACAGTCACCAATACAGACTTCAGTA 11

CG038009 973 bp DNA linear GSS 19-AUG-2003  
PUF0241TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMBTA0713G10,  
genomic survey sequence.

ACCESSION CG038009  
VERSION CG038009.1 GI:33910165  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 973)  
Whiteclaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and  
Bennetzen, J.  
Maize Genomics Consortium  
Unpublished (2003)  
Other GSSs: PUF0241TB  
Contact: Cathy Whitelaw  
TIGR Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: sheared ends.

FEATURES  
source  
1. 973  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone\_lib="ZMMBTA0713G10"  
/clone="ZM 0.6 1.0 KB"  
/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
CoT selected genomic DNA library"

ORIGIN  
Query Match 3.6%; Score 39; DB 29; Length 973;  
Best Local Similarity 48.8%; Pred. No. 3.1; Indels 0; Gaps 0;  
Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 579 AGTACGAATCTGTATGATAATCAATGGTAGCAGTGTGTAACACGCGGATCGGATCTA 638  
Db 397 AGTAACAATATGTTACATATATATTTTAAAGTCTGTAATCAAGGCTACTTTTGA 456  
QY 639 GCACCTGATGTCTGGTATGTAATCTGCTATGTCATAGCGGCACATAGAAAGAT 698  
Db 457 GCACCAATATTTTGGAACTAATTCACATCATTTTAAATTAGACAAACAGAGCTAT 516  
QY 699 ACCTCAGTCTACGATACGTAAGTCTGTATACATTTCACTGCGAAATTTCTGAACAAATTGGA 758

Db 517 AACTCAGTTCAGAGTTCGAATGCTCTCTTTTAAAGTCTCATCAGCGCTGACATTAAGA 576  
QY 759 GAGCAATATGAAATCTAAATGAAGAACTCTCTCATTA 793  
Db 577 AAGAGGTCCACAGTTCAGCAAAATACAGACTTCAGTA 611

RESULT 11  
LOCUS CB943651  
DEFINITION AGENCOURT\_138622618 NICHD\_XGC\_Tad2 Xenopus laevis cDNA clone  
IMAGE:6933934 5', mRNA sequence.  
ACCESSION CB943651  
VERSION CB943651.1 GI:30191094  
KEYWORDS EST.  
SOURCE Xenopus laevis (African clawed frog)  
ORGANISM Xenopus laevis

REFERENCE 1 (bases 1 to 505)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapsb-remail.nih.gov  
Tissue Procurement: Drs. Donald Brown and Liqian Cai  
cDNA Library Preparation: CLONTECH  
DNA Sequencing by: Agencourt Bioscience Corporation  
Cloning through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LICM3249 row: 0 column: 21  
High quality sequence stop: 459.

FEATURES  
source  
1. 505  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/cnames="IMAGE:6933934"  
/dev\_stage="metamorphosis stage 62"  
/clone\_lib="NICHD XGC Tad2"  
/note="Organ: Developing Tadpole; Vector: pDNR-LIB;  
Site 1: SfiI; Site 2: SfiI; 5' and 3' adaptors were used  
in cloning as follows: 5' adaptor sequence:  
5'-CATCGCCATTATGGCC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCGCGGCGGCGCATG-dT(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.7 kb  
(range 0.8-3.0 kb). 15/15 colonies contained inserts by  
PCR. This library was enriched for full-length clones and  
was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN  
Query Match 3.6%; Score 38.8; DB 14; Length 505;  
Best Local Similarity 56.2%; Pred. No. 2.4;  
Matches 73; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 683 GCGACATAGAAAGTACCTCAGTGTCTCAGCATCAGTTCAGCTCTGATCTTCACTGCAAA 742  
Db 351 GCACATATACAGAGTACCTCAAGTCAAGAAACAGACCTGACGCTACTGTACG 410  
QY 743 TTTCTGAACAATTGGAGAGCATTTATGAATACTAAATGGAACTCTCTCAATTAAAGTGA 802  
Db 411 TTTGTTTACAGCCATCATATTTATTACTTTCTTAATATAAACACATCGCGTTAAGTTCAA 470  
QY 803 AACAGAGCGC 812  
Db 471 AAAAGACCAC 480



REFERENCE 1 (bases 1 to 823)  
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.M., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A., and Hubbard, S.J.  
 A Comprehensive Collection of Chicken cDNAs  
 Curt. Biol. 12 (22), 1965-1969 (2002)  
 22335534  
 12445392  
 Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk  
 Location/Qualifiers  
 1..823  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Compton Line 151"  
 /db\_xref="taxon:9031"  
 /clone="ChST547C20"  
 /sex="female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="CSQCHN74"  
 /note="Organ: kidney + adrenal; Vector: pBluescript II  
 KS(+); Site 1: EcoRI; Site 2: NotI. This normalized  
 library was constructed from 1 million independent clones.  
 cDNA synthesis was initiated using an oligo(dT) primer,  
 using methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was blunt-ended, ligated to NotI adapters, digested with  
 EcoRI, size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pBluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS  
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."

ORIGIN  
 Query Match 3.5%; Score 38.2; DB 13; Length 823;  
 Best Local Similarity 51.5%; Pred. No. 4.9;  
 Matches 88; Conservative 0; Mismatches 83; Indels 0; Gaps 0;  
 QY 562 AGAAGGTGACGGGTAGTACGAATCTGTATGATTAATCAATGAGTACGATGATGATAA 621  
 DB 760 ACATGGCTCTTGTGTTTATAGTCTATCAGGAGATAGGATGGGGCAAAAAGCAAG 701  
 QY 622 CAGCGATCGGATCTAGCAGTCTGTCTGGGTATGTATCTCTGGCTATCTTCTCATAG 681  
 DB 700 CAGCAAAAGATGCACAGGAACCTTGAACATTTGCCAAGAATTCATGTTCTTCAGAAC 641  
 QY 682 GCGACATAGAAAGATACCTCAGTGTGAGTACGATGATGATGATGATGATGATGAT 732  
 DB 640 TGCATGTAGAAATTAACCTGATGATGATGATGATGATGATGATGATGATGATGAT 590

RESULT 14  
 CG865000 839 bp DNA linear GSS 19-NOV-2003  
 ZMBB0273K14f ZMBB0273K14 5', genomic survey sequence.  
 ZMBB0273K14 5', genomic survey sequence.  
 CG865000  
 CG865000.1 GI:38441215  
 GSS.  
 Zea mays subsp. mays (maize)  
 Zea mays subsp. mays  
 Zea mays subsp. mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 839)

RESULT 12  
 CE323888 492 bp DNA linear GSS 26-SEP-2003  
 LOCUS tigr-gss-dog-17000360301767 Dog Library Canis familiaris genomic,  
 DEFINITION genomic survey sequence.  
 ACCESSION CE323888  
 VERSION CE323888.1 GI:36133243  
 KEYWORDS GSS.  
 SOURCE Canis familiaris (dog)  
 ORGANISM Canis familiaris  
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Eukaryota; Metazoa; Carnivora; Fissipedia; Canidae; Canis.  
 1 (bases 1 to 492)  
 Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,  
 Rusch, D.B., Deicher, A.L., Pop, M., Wang, W., Fraser, C.M., and  
 Venter, J.C.  
 The dog genome: survey sequencing and comparative analysis  
 Science 301 (5641), 1898-1903 (2003)  
 22875432  
 14512627  
 Contact: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
 Rockville, MD 20850, USA  
 Tel: 301-838-0200  
 Fax: 301-838-0208  
 Email: ekirknes@tigr.org  
 Class: shotgun.  
 Location/Qualifiers  
 1..492  
 /organism="Canis familiaris"  
 /mol\_type="genomic DNA"  
 /strain="Standard Poodle"  
 /db\_xref="taxon:9615"  
 /clone\_lib="Dog Library"  
 /note="Site 1: BstXI; Libraries were prepared from  
 peripheral blood"

ORIGIN  
 Query Match 3.5%; Score 38.4; DB 29; Length 492;  
 Best Local Similarity 47.5%; Pred. No. 3.2;  
 Matches 114; Conservative 0; Mismatches 126; Indels 0; Gaps 0;  
 QY 669 CTATGTTCTAAGGGGACATAGAAAGATACCTCAGTGTGACATACGTAAGCTCTGTA 728  
 DB 130 CTGCTTTTATAATGGGAATAGACAAGTATCTCTGTGCTATTATCCCTCCACATT 189  
 QY 729 CATTCTACTGCAATTTCTGCAAAATGGAGCATATGAAATCTAATGGAATCTCT 788  
 DB 190 GAATCAGCGTCCCTTCCATACAGTAAATAGAAAGAAATACATATGTAAACCA 249  
 QY 789 CATTATAAGTGGAAACAGAGCGCCCTTTTATTATGAAACAGAGCGGTCAAGACGCTTT 848  
 DB 250 CAAAAGAGAAAGAAAGAAATGTCATTTTACAGAAATATATTTGTGAGCAAGCAGA 309  
 QY 849 TCACTCATCAGAGCGGTTCATCCAGATCATCTTCCCTTGAACCATGTTCTGCGAT 908  
 DB 310 AAAGCAAAATATTAAGATTTAAAGACGGCCCCCTGGGGTGGCTCAGTGGTTGACAT 369

RESULT 13  
 BU374247/c 823 bp mRNA linear EST 28-NOV-2002  
 LOCUS BU374247 823 bp mRNA linear EST 28-NOV-2002  
 DEFINITION 603588074f1 CSQCHN74 Gallus gallus cDNA clone ChST547C20 5', mRNA  
 sequence.  
 ACCESSION BU374247  
 VERSION BU374247.1 GI:25982248  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.

**AUTHORS** Bharti, A.K., Young, S., Kavchok, S., Keizer, G., Bronzino, A.C.,  
Zohovetz, V., Fuks, G., Yu, Y., Wing, R. and Messing, J.  
**TITLE** Sequencing of the maize genome at PGIR (2003c)  
**JOURNAL** Unpublished (2003)  
**COMMENT** Contact: Bharti, A.K.

Dr. Joachim Messing's lab  
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers  
University  
190 Frelinghuysen Road, Piscataway, NJ 08854, USA  
Tel: 732 445 3801  
Fax: 732 445 5735  
Email: bharti@waksman.rutgers.edu  
Seq primer: T7  
Class: BAC ends  
High quality sequence start: 507.

**FEATURES**

source  
1. .839  
Location/Qualifiers  
/organism="Zea mays subsp. mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/sub\_species="mays"  
/db\_xref="taxon:4578"  
/clone="ZMBBc0273K14"  
/lab\_host="E. coli DH10B"  
/clone\_lib="ZMBBc (EcoRI)"  
/note="Vector: pTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI"

**ORIGIN**

Query Match 3.5%; Score 38.2; DB 29; Length 839;  
Best Local Similarity 50.3%; Pred. No. 5;  
Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
QY 632 GCATCTAGCACTGCTATGCTGGGTATGTAATCCTGGCTATGTTCTAAGGGCGCATAG 691  
DB 815 GCATCATGTGGGTTTACTTGCTGCGCATTTGGTATATAAATAATGAGATATAA 756  
QY 692 AAGAATACCTCAGTGTACGATAGCACTCTGTACATTTTCACGCAAAATTTCTGAAC 751  
DB 755 ACAATATATTAAATGACATCTAGAATTTAACTTCTACATTTGAAACCAAACTCTAAAG 696  
QY 752 AATTGGAGACATTTGAATATCTAAATGGAATCTCTCATTTAAGTGAACACAGAGCG 811  
DB 695 AAAGAGAGAGGGAAGGAATCTTAAATATATAGTCCAGAGAAAGTTTATATATATAT 636  
QY 812 CCTTTT 818  
DB 635 CCCTAT 629

**RESULT 15**

AW384856  
LOCUS RC3-HT0419-301299-011-d02 HT0419 Homo sapiens cDNA, mRNA sequence.  
DEFINITION RC3-HT0419-301299-011-d02 HT0419 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW384856  
VERSION AW384856.1 GI:6889515  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

**ORGANISM**

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE**

1 (bases 1 to 586)  
HCGP <http://www.ludwig.org.br/ORESTES>.  
The FAPESP/LiCR Human Cancer Genome Project

**AUTHORS**

Unpublished (1999)  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

**JOURNAL**

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LiCR Human Cancer Genome

Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC3&t2=RC3-HT0419-301299-011-d02&t3=1999-12-30&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 9  
High quality sequence stop: 585.

**FEATURES**

source  
1. .586  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="HT0419"  
/note="Organ: head neck; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

**ORIGIN**

Query Match 3.5%; Score 37.8; DB 10; Length 586;  
Best Local Similarity 54.7%; Pred. No. 5.4;  
Matches 75; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 513 GACAGTGTCTGCTGTTGAACATGCTCATGTCTATATGAATGAATAAGAAAGGTGCA 572  
DB 99 GACAGTGTCTGCTGTTGAACATGCTCATGTCTATATGAATGAATAAGAAAGGTGCA 158  
QY 573 CGGGTTAGTACGAATCTGTATGATATCAATGGTACGATGATGTTAAACACCGGATCGG 632  
DB 159 TGGATTAAATAAGAGAGCTTCATGCCCTACAAAGAACTGATACACAGTAAATCAGGAAGGGG 218  
QY 633 GATCTAGCAGCTGCTATG 649  
DB 219 GATCAAGAACTCTCTATG 235

Search completed: June 10, 2004, 13:10:48  
Job time : 2144.6 secs

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 08:08:35 ; Search time 332.836 Seconds  
(without alignments)  
13899.608 Million cell updates/sec

Title: US-10-070-386-2  
Perfect score: 1089  
Sequence: 1 taactccatcatatagcggg.....actgcaatagtgccgcggg 1089

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1089	100.0	1089	5	Aaf79835 Mycelia s
2	47	4.3	2000	7	Ada71938 Rice gene
3	44.8	4.1	2000	7	Ada71938 Rice gene
4	44	4.0	14041	4	Aah48024 Internal
5	41.2	3.8	10732	3	Aaa10594 Gene enco
6	40.8	3.7	276	9	Add17664 DNA (Seqi
7	36.2	3.3	4998	2	Aax23518 Human kid
8	36.2	3.3	177380	7	Acf62751 Cancer ba
9	36.2	3.3	177380	7	Adb20870 MRP1 base
10	36.2	3.3	177380	9	Adb87959 Human UGT
11	36.2	3.3	177380	9	Adb96942 Human MDR
12	36.2	3.3	177380	9	Adb92133 Human MDR
13	35.4	3.3	7025	6	Aax40059 Human che
14	35.4	3.3	7025	6	Aax63350 Chemical
15	35.2	3.2	2604	4	Aah16164 Human cdn
16	35.2	3.2	2604	6	Aht07177 Human ova
17	35.2	3.2	4020	4	Aal06054 Human rep
18	35.2	3.2	4020	4	Abl98619 Human tes
19	35	3.2	4072	4	Abl12544 Drosophil
20	34.8	3.2	40308	4	Aax68184 Human imm
21	34.4	3.2	1370	6	Abq69040 Listeria
22	34.4	3.2	5818	6	Abq71014 Listeria
23	34.4	3.2	110000	6	Abq67196_4

Continuation (5 of

C 24	34.4	3.2	110000	6	ABQ69245_27	Continuation (28 o
C 25	34.4	3.2	143306	6	ABK49586	Abk49586 Human tra
C 26	34.2	3.1	576	4	AAI16818	AAI16818 Probe #67
C 27	34.2	3.1	576	4	ABA60543	ABA60543 Human foe
C 28	34.2	3.1	576	4	AAI40432	AAI40432 Probe #91
C 29	34.2	3.1	576	4	ABA28703	ABA28703 Probe #71
C 30	34.2	3.1	576	4	AAK34711	AAK34711 Human bon
C 31	34.2	3.1	576	4	AAK08823	AAK08823 Human bra
C 32	34.2	3.1	576	4	ABS34484	ABS34484 Human liv
C 33	34.2	3.1	576	6	ABS09300	ABS09300 Human imm
C 34	34.2	3.1	5835	4	AAK79610	AAK79610 Human imm
C 35	34.2	3.1	5835	4	AAK72786	AAK72786 Human imm
C 36	34.2	3.1	5835	4	AAK79611	AAK79611 Human imm
C 37	34	3.1	5449	5	AAS85687	AAS85687 DNA encod
C 38	33.8	3.1	366	7	ACA22237	ACA22237 Prokaryot
C 39	33.6	3.1	432	7	ABX53039	ABX53039 Bovine ES
C 40	33.4	3.1	2853	7	ACA23231	ACA23231 Prokaryot
C 41	33.4	3.1	110000	2	AAK20248_00	AAK20248 Borrelia
C 42	33.2	3.0	562	4	AAK37880	AAK37880 Human bon
C 43	33.2	3.0	1048	6	ABN74166	ABN74166 Bovine em
C 44	33	3.0	1957	6	ABS55349	ABS55349 cDNA enco
C 45	33	3.0	19736	6	ABA92158	ABA92158 Human pro

ALIGNMENTS

RESULT 1  
AAF79835  
ID AAF79835 standard; DNA; 1089 BP.  
XX  
AC AAF79835;  
DT 30-MAY-2001 (first entry)  
XX  
DE Mycelia sterilia terminator sequence.  
XX  
KW Promoter; terminator; regulatory region; filamentous fungus;  
KW protein production; ds.  
XX  
OS Mycelia sterilia.  
XX  
PN WO200118219-A1.  
PD 15-MAR-2001.  
XX  
PF 07-SEP-2000; 2000WO-JP006104.  
XX  
PR 07-SEP-1999; 99JP-00252851.  
XX  
PA (MEIJ ) MEIJI SEIKA KAISHA LTD.  
XX  
PI Watanabe M, Murakami T;  
XX  
WPI; 2001-235202/24.  
XX  
PT New promoter and terminator functioning synchronously, useful for  
PT regulating expression of endogenous gene in filamentous fungi,  
PT particularly transformed Mycelia sterilia capable of producing high  
PT yields of target protein/substance.  
XX  
PS Claim 6; Page 20-22; 26pp; Japanese.  
XX  
CC The present invention provides the sequences of a promoter and terminator  
CC from Mycelia sterilia which can be used to regulate the expression of an  
CC endogenous gene in filamentous fungi, particularly transformed Mycelia  
CC sterilia, to produce useful target proteins or other substances. The  
CC present sequence is the terminator of the invention  
XX  
SQ Sequence 1089 BP; 298 A; 248 C; 247 G; 296 T; 0 U; 0 Other;  
Query Match 100.0%; Score 1089; DB 5; Length 1089;  
Best Local Similarity 100.0%; Pred. No. 0;



```

Db 849 YTMVTCYAMTCAKCYKAMTKMTWACAWRATSWRWRAMGVRKXKRYKRAVWNRW 790
Qy 482 CACTACTAGAACACGACAGTGTGAATCCGACAGTGTCTGCTGTGTGAACATGTCTAT 541
Db 789 HRCWAGARWKSRYRKKYATRYKWMAMTWMSWRKWSYRMSGKGRWRSAN 730
Qy 542 GTCTATGATGAATAAGAGAGAGTGTGACGGGTAGTACCAATCTGTATGATAATCA 601
Db 729 RYCSRMCAKTKYASSARWTGAKESYRVRWYWKRGWTYRYRWSRMTARMKSR 670
Qy 602 ATGTAGACAGTATGGTAACA----GGGATCGGATCTAGCACTGTATGTCTGGGTA 657
Db 669 RHWAGASMKCMYWRGARSMMYSKYSKSAKCKKTRTYSMTGMYGYSYKSMWS 610
Qy 658 TGTATCTCGCTATGTTTCATAAGCGGACATAGAAAGAAATACCTCAGTGTGAGATACG 717
Db 609 TSXNSYMGKMTCTMTYSKMGSTRSKMGWSGMSRMYRWKMKOMKRYMYMKWKTWR 550
Qy 718 TAAGCTCTGTACATTTCTGCAAAATTTCTGAACAAATTTGGAGAGCAATTAAGAAATACTAA 777
Db 549 RCMCYRWGYTMYTTSRSMYTGKARYTSKRYMYKRYKRYCHYIYGYMYMKSYMYR 490
Qy 778 ATGGAACCTCATTATAGTGAAGAAACAGAGCGCCCTTTATTATGAAACAGAGAGTC 837
Db 489 YGYCKKCKCCYAMCKAAYSEMMYWKYKSKWRMSTKTNWSMWYKCKRSMYKAGCC 430
Qy 838 AAGAACGCTTTCAACGTCATCAGAGCGGTTCCATCCAGATCATACTTTCCTTGAACA 897
Db 429 YGCRMTYCSYGMKWTYMGSYKYSRCYKRYMYRMVYKMMYMYAYSSNMWTWYTYA 370
Qy 898 TGTCTCGCATTCAGAACTGATGAGATGGAACCGTCCAGGGTTCCTGTCTATCCCTTG 957
Db 359 KYMKYWKRRGTMSYKYSKYYCTWYKMKRCYRWKRMKRYKRYKRYKRYKRYKRYKRY 310
Qy 958 GTCCTCTGCAATAAAATCGTATTACCA-TTTTCTTTCGACGCGCGGTCAACGTGAGCG 1016
Db 309 CYWCCYKRGWYSRSMWRTAGKWMRSWCRYSYKRYMYKMMKYSYMSVGVARS 250
Qy 1017 ACGTGCCACGTTGAGTCCCAATGACAGCATGATGATGATGATGATGATGATGATGATG 1065
Db 249 SGTWRSAAKRYTGYSTRSRAKWMACRMYSACRRYSRTSYCGCSYC 201

```

## RESULT 3

```

ID ADA71938 standard; DNA; 2000 BP.
AC ADA71938;
DT 20-NOV-2003 (first entry)
DE Rice gene, SEQ ID 5263.
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
OS Oryza sativa.
PN WO2003000898-A1.
PD 03-JAN-2003.
PF 22-JUN-2001; 2001WO-IB001105.
PR 22-JUN-2001; 2001WO-IB001105.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,
PI Katagiri F, Quan Y, Tao Y, Whitham S, Xie Z, Zhu T, Zou G,
XX WPI; 2003-175290/17.
DR

```

```

XX Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
XX gene expression.

```

Claim 27; SEQ ID NO 5263; 899pp; English.

The present invention relates to a method (M1) for identifying genes involved in plant resistance or response to pathogenic infection. M1 comprises identifying a gene whose expression is significantly altered in the incompatible interaction of plant gene expression relative to the expression of the gene in an uninfected plant, in a mutant plant that does not express a gene associated with response to pathogenic infection, or in a corresponding incompatible or compatible interaction. (M1) is useful for conferring resistance to resistance or tolerance to a plant to bacterial, fungal or viral infection. The present sequence was used to illustrate the invention.

Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 4.1%; Score 44.8; DB 7; Length 2000;

Best Local Similarity 11.3%; Pred. No. 0.0024;

Matches 104; Conservative 398; Mismatches 410; Indels 12; Gaps 4;

```

Qy 18 CGTTGGCTGAAGAGGAGAGCGGACAGAGATCATCTTTTACACGAGTTGAGATCT 77
Db 73 SGTGSKWSGSGYSGKMKRYKESKWRGRGRGRMRMRMRMRMRMRMRMRMRMRMR 132
Qy 78 GATCTGTGTTTCACTCTCAGTAACCTGTGTGGAAGTTACACTTCTGCTCTCCCTCTAC 137
Db 133 GRWVGSRMSYMWYVARGCGSKKSKGSGWKTCTRGAGSGWSGSKYKSGMS 192
Qy 138 CAGCCTTCAGGCAACCAAGATGTTAGAGTTTCGTCATTTATATAGCTCTGGCGA 197
Db 193 KRWMSGSGRSGRRSRYTGTSRKYTKYKMYTYSASRCWRAITWITSYMACSYT 252
Qy 198 TGACTACATTTAGGCGATGACGACATGCTCTACTGCTCTGCTGCTGTTGGTTAGGTA 257
Db 253 WCSRKSRMWRKMRKMRKMRKMRKMRKMRKMRKMRKMRKMRKMRKMRKMRKMRKMR 312
Qy 258 CCTAGCTAGCAATATCAACATACAAATGTGGTTTCCACATGTCTAGCTGGTTCTACC 317
Db 313 YWGRGYSRMVYKMYVRYGKMKRWGAGRMWRMRMRMRMRMRMRMRMRMRMRMR 372
Qy 318 TAGTCTGAGTGAATGGTAAATTTGATATATTGAGCTTGACCCCGCAATATTGTAACAGAG 377
Db 373 RWAQKSSRTSRKRRKMRKMRKMRKMRKMRKMRKMRKMRKMRKMRKMRKMRKMRKMR 432
Qy 378 CCAACAATGGCTCACCTGGCCCCCAGACATGTGGGTATATAAGCTACCTGTCTA----- 432
Db 433 TCRMKSYGMRMRKSKMRMSKYSKMSRYMRKMRKMRKMRKMRKMRKMRKMRKMRKMR 492
Qy 433 --GCAATCAGACTTACTAGTAAAGCGTGTAGTACGAATCTGTATGATAAATAGTCACTAG 490
Db 493 RSGMRKCRRRRWRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMR 552
Qy 491 AACTACGACAGTGTGAATCCGACAGTGTCTGCTGTGGAACATGTCTATGTCTATATG 550
Db 553 GMMKRYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYK 612
Qy 551 AATGAATAAGAGAGAGTGTGACGGGTGTAGTACGAATCTGTATGATAAATAGTCACTAGCA 610
Db 613 SMRSRKRKCRKCRKCRKCRKCRKCRKCRKCRKCRKCRKCRKCRKCRKCRKCRKCR 671
Qy 611 GTGATGTAACACAGCGGATCGGATCTAGCACCTGCTATGTCTGGGTATGTAATCCTGGCT 670
Db 672 SKTYAKYGSYWRYYRANWYRANWYRANWYRANWYRANWYRANWYRANWYRANWYR 731
Qy 671 ATGTTCAAGGCGCATAGAAAGAAATACCTCAGTGTGAGCATAGCTAGCTCTGTA-C 729
Db 732 SWYKCYKSKYKSMYMYNMAKTKWKRVRATRRMMWYRYSKMYKWTCTWGTWGYWY 791

```







```
XX OS Homo sapiens.
XX KW cytochrome p450; subfamily IIIA; nifedipine oxidase; polypeptide 5;
XX KW cytosolic; gene; ds.
XX PD 11-MAR-1999.
XX PF 02-SEP-1998; 98WO-US018426.
XX PR 02-SEP-1997; 97US-0057854P.
XX PA (MEDI-) MEDICAL COLLEGE GEORGIA RES INST.
XX PI Ryan JW, Sprinkle TJC, Venema RC;
XX WPI; 1999-205193/17.
XX Nucleic acid encoding human aminopeptidase P.
XX Claim 13; Page 109-139; 201pp; English.
XX PS
XX CC This invention describes the isolation of a novel human aminopeptidase P
XX CC (Amp). This protein is used to produce recombinant Amp and can be used
XX CC for gene therapy for treating Amp-deficiency conditions. Its fragments
XX CC are used as primers and probes to identify patients with homozygous and
XX CC heterozygous Amp deficiency, including prenatal diagnosis (patients
XX CC defective in Amp are at risk of developing angioedema if treated with
XX CC angiotensin-converting enzyme inhibitors), also as antisense inhibitors
XX CC in cases of excessive Amp expression. The product of the invention is
XX CC also used to identify Amp-expressing sequences in other animals and to
XX CC generate transgenic animals, and comparisons of genomic sequences are
XX CC used to detect mutations. Amp inhibitors are potentially useful as
XX CC antihypertensive agents and to prevent or treat arterial (re)stenosis or
XX CC atherosclerosis. The structure of Amp is used to design synthetic
XX CC substrates, e.g. for use in Amp assays. Amp, which hydrolyzes N-terminal
XX CC imido bonds, can be used to degrade industrial protein feeds to free
XX CC amino acids, to degrade proteinaceous wastes, as additives in enzyme
XX CC formulations used to treat malabsorption syndrome and for studying its
XX CC biological role. Antibodies against Amp are used in immunohistochemical
XX CC methods to study Amp distribution.
XX SQ Sequence 49998 BP; 12605 A; 11725 C; 11351 G; 14317 T; 0 U; 0 Other;
Query Match 3.3%; Score 36.2; DB 2; Length 49998;
Best Local Similarity 48.8%; Pred. No. 8.8;
Matches 98; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 380 AACATGGCTCAGCTGGCCGCCGAGACATGTGGCTATATAGCTACCTCTCTAGCAATCA 439
DB 23385 AAAATGGACTAAATATTTGAACAGACAAATTCACCAATAAGATATCTGGATGGCAATA 23326
QY 440 GACTTACTGATAGAACGTCCTCCCTATATGTCTATAAATTAAGTCACCTACTAGAACTACCGA 499
DB 23325 AACACATATAAGATGCTTTCCTCATTTAGCCATTAGGAGATATAAAATTAACACAGT 23266
QY 500 CAGTGTGAATCCGACAGTGTCTGGTCTGTGACATGTCATGTCATATGATGATGATGA 559
DB 23265 GAGATACCACTACAGATCTATTTCGAATGGCTAAAGTTAAAAAGACTGACTGAAGTACTAT 23206
QY 560 GAAGAAGGTGTACGGGTTAG 580
DB 23205 TCCAAATTAGTGAGGTAGTAG 23185
RESULT 8
ACF62751/c
ID ACF62751 standard; DNA; 177380 BP.
XX AC ACF62751;
XX AC
XX 08-OCT-2003 (first entry)
DT
XX
XX DE Cancer based on CYP3A5 related polynucleotide SEQ ID NO:683.
```

```
XX Cancer; CYP3A5; irinotecan; pharmaceutical; malignant glioma;
XX KW cytochrome p450; subfamily IIIA; nifedipine oxidase; polypeptide 5;
XX KW cytosolic; gene; ds.
XX OS Unidentified.
XX PN WO2003013534-A2.
XX PD 20-FEB-2003.
XX PF 23-JUL-2002; 2002WO-EP008219.
XX PR 23-JUL-2001; 2001EP-00117608.
XX PR 24-MAY-2002; 2002EP-00011710.
XX PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX PI Heinrich G, Kerb R;
XX WPI; 2003-268144/26.
XX DR
XX CC New use of irinotecan for preparation of compositions for treating cancer
XX CC in subject having genome with variant allele comprising cytochrome p450,
XX CC subfamily IIIA, polypeptide 5 polynucleotide, termed CYP3A5.
XX PS Disclosure; SEQ ID NO 683; 86pp; English.
XX CC The present invention describes the use of irinotecan (I) or its
XX CC derivative for the preparation of a pharmaceutical composition for
XX CC treating colorectal, cervical, gastric, lung, ovarian or pancreatic
XX CC cancer, or malignant glioma in a subject having a genome with a variant
XX CC allele which comprises a cytochrome p450, subfamily IIIA (nifedipine
XX CC oxidase), polypeptide 5 (CYP3A5) polynucleotide (II). (I) and (II) have
XX CC cytostatic activity. The therapeutic applications of (I) is improved,
XX CC since it is possible to individually treat a subject with an appropriate
XX CC dosage and/or an appropriate derivative of (I). Therefore, undesirable,
XX CC harmful or toxic effects are efficiently avoided. Unnecessary and
XX CC potentially harmful treatment of those subjects who do not respond to the
XX CC treatment with substances (nonresponders), as well as the development of
XX CC drug resistances due to suboptimal drug dosing can be avoided. ACF62200
XX CC to ACF62751 and ABW34912 to ABW35013 represent sequences used in the
XX CC exemplification of the present invention
XX SQ Sequence 177380 BP; 58338 A; 32530 C; 31389 G; 55123 T; 0 U; 0 Other;
Query Match 3.3%; Score 36.2; DB 7; Length 177380;
Best Local Similarity 50.9%; Pred. No. 17;
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 182 TATATGGCTCTGGCGATGAGTAGCATTTATGAGCGATGCACGATGCTCTACTGTGC 241
DB 47048 TTTATGTGTAGCTCTCTGTGTGGACTTCAGCCGTCATGTACGATTTTCATCCACGATAC 46989
QY 242 TCTGTGTGTTAGGTTACCTTAGCTAGACAAATATCACATACAAAATGTGTTTCCACATG 301
DB 46988 TTTAAGCCTTATGTTTATATATGCTGTGGAGGAGGAAAACAAAACGTTTCTTCCATT 46929
QY 302 TCAGCTGTTCTACCGTAGTCTGAGTGAATGGGTAATGTATATATGA 350
DB 46928 TCAGAGAGAGTAAATTAATCTGGGATAAAAAAGCCCTATTAAATTGATTCA 46880
RESULT 9
ADB20870/c
ID ADB20870 standard; DNA; 177380 BP.
XX AC ADB20870;
XX AC
XX 20-NOV-2003 (first entry)
DT
XX
XX DE MRPI based cancer related nucleic acid SEQ ID NO:683.
```



XX WO2003013537-A2.  
XX 20-FEB-2003.  
XX 23-JUL-2002; 2002WO-EP008218.  
XX 23-JUL-2001; 2001EP-00117608.  
XX 24-MAY-2002; 2002EP-00011710.  
XX (EPID-) EPIDAUS BIOTECHNOLOGIE AG.  
XX Heinrich G, Kerb R;  
XX WPI; 2003-368145/36.  
XX New use of irinotecan for preparation of pharmaceutical compositions for  
XX treating cancer in subject having genome with variant allele comprising  
XX multidrug resistance 1 polynucleotide.  
XX Disclosure; SEQ ID NO 683; 130pp; English.  
XX The invention relates to the novel use of irinotecan or its derivative  
XX for the preparation of pharmaceutical compositions for treating  
XX colorectal, cervical, gastric, lung, ovarian or pancreatic cancer, or  
XX malignant glioma in a subject having a genome with a variant allele which  
XX comprises a multidrug resistance 1 (MDR1) polynucleotide. A composition  
XX of the invention has cytostatic activity. The invention is useful for the  
XX preparation of pharmaceutical compositions for treating colorectal,  
XX cervical, gastric, lung, ovarian or pancreatic cancer, or malignant  
XX glioma in a subject (preferably human, more preferably African or Asian)  
XX or a mouse. The present sequence is used in the exemplification of the  
XX invention.  
XX  
XX Sequence 177380 BP; 58338 A; 32530 C; 31389 G; 55123 T; 0 U; 0 Other;  
XX  
XX Query Match 3.3%; Score 36.2; DB 9; Length 177380;  
XX Best Local Similarity 50.9%; Pred. No. 17;  
XX Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;  
XX  
XX QY 182 TATATGGCTCTGGCGATGATGACATTTATGAGGATGCACGATGCTTCTACTGCTGC 241  
XX DB 47048 TTTATGTTAGTCTCTCTGTGGACTTCAGCGCTCATGTACGATTTTCATTCACCGATAC 46989  
XX  
XX QY 242 TCTGTGGTTAGTTACCTTACCTAGACATATCAATACAAATGTTTCCACATG 301  
XX DB 46988 TTTAAGCCTTATGTTAATATGCTGTGGAGGAAAAACAAACGTTTGTCTCCATT 46929  
XX  
XX QY 302 TCAGCTGGTTCTACCGTAGTCTGAGTGAAATGGTAAATTGATATATGA 350  
XX DB 46928 TCAGAGAGAGTAAATTAATCTGGGATATAAAAGCCCTATTAAATTGATTCA 46880  
XX  
XX  
XX RESULT 12  
XX ADB92133/C  
XX ID ADB92133 standard; DNA; 177380 BP.  
XX AC ADB92133;  
XX  
XX DT 04-DEC-2003 (first entry)  
XX  
XX DE Human MDR1 related DNA sequence SEQ ID NO:683.  
XX  
XX KW irinotecan; colorectal cancer; cervical cancer; gastric cancer;  
XX lung cancer; ovarian cancer; pancreatic cancer; malignant glioma;  
XX multidrug resistance 1; MDR1; cytostatic; human; UGT1A1; MRP1; TOPI; ds.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO2003013535-A2.  
XX  
XX PD 20-FEB-2003.  
XX

XX 23-JUL-2002; 2002WO-EP008220.  
XX 23-JUL-2001; 2001EP-00117608.  
XX 24-MAY-2002; 2002EP-00011710.  
XX (EPID-) EPIDAUS BIOTECHNOLOGIE AG.  
XX Heinrich G, Kerb R;  
XX WPI; 2003-342400/32.  
XX  
XX New use of irinotecan for preparation of pharmaceutical compositions for  
XX treating cancer in subject having genome with variant allele comprising  
XX multidrug resistance 1 polynucleotide.  
XX Disclosure; SEQ ID NO 683; 104pp; English.  
XX The invention relates to a novel use of irinotecan or its derivative for  
XX the preparation of a pharmaceutical composition for treating colorectal,  
XX cervical, gastric, lung, ovarian or pancreatic cancer, or malignant  
XX glioma in a subject having a genome with a variant allele which comprises  
XX a multidrug resistance 1 (MDR1) polynucleotide. A composition of the  
XX invention has cytostatic activity. The present sequence is used in the  
XX exemplification of the invention.  
XX  
XX Sequence 177380 BP; 58338 A; 32530 C; 31389 G; 55123 T; 0 U; 0 Other;  
XX  
XX Query Match 3.3%; Score 36.2; DB 9; Length 177380;  
XX Best Local Similarity 50.9%; Pred. No. 17;  
XX Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;  
XX  
XX QY 182 TATATGGCTCTGGCGATGATGACATTTATGAGGATGCACGATGCTTCTACTGCTGC 241  
XX DB 47048 TTTATGTTAGTCTCTCTGTGGACTTCAGCGCTCATGTACGATTTTCATTCACCGATAC 46989  
XX  
XX QY 242 TCTGTGGTTAGTTACCTTACCTAGACATATCAATACAAATGTTTCCACATG 301  
XX DB 46988 TTTAAGCCTTATGTTAATATGCTGTGGAGGAAAAACAAACGTTTGTCTCCATT 46929  
XX  
XX QY 302 TCAGCTGGTTCTACCGTAGTCTGAGTGAAATGGTAAATTGATATATGA 350  
XX DB 46928 TCAGAGAGAGTAAATTAATCTGGGATATAAAAGCCCTATTAAATTGATTCA 46880  
XX  
XX  
XX RESULT 13  
XX ABR40059  
XX ID ABR40059 standard; DNA; 7025 BP.  
XX  
XX AC ABR40059;  
XX  
XX DT 21-MAY-2002 (first entry)  
XX  
XX DE Human chemically pretreated gene sequence #71 strand 1.  
XX  
XX KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;  
XX cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;  
XX UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200202806-A2.  
XX  
XX PD 10-JAN-2002.  
XX  
XX PF 29-JUN-2001; 2001WO-EP007470.  
XX  
XX PR 30-JUN-2000; 2000DE-01032529.  
XX  
XX PR 01-SEP-2000; 2000DE-01043826.  
XX  
XX PA (EPID-) EPIDENOMICS AG.  
XX  
XX PI Olek A, Piepenbrock C, Berlin K;  
XX

DR WPI; 2002-154757/20.  
 XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligonucleotides,  
 PT useful for detecting cytosine methylation state of genes associated with  
 PT pharmacogenomics and for therapy of diseases e.g. cancer.  
 XX  
 PS Claim 1; SEQ ID NO 141; 24pp; English.  
 XX  
 CC The invention relates to a nucleic acid comprising a sequence at least 18  
 CC bases in length of a segment of the chemically pretreated DNA of genes  
 CC associated with pharmacogenomics according to one of the sequences of the  
 CC genes ALDH6 (NM\_000693), CYP11A (NM\_000781), CYP11B1 (NM\_000497), CYP3A3  
 CC (NM\_000776 and NM\_017460), DPYD (NM\_000110), EPHX2 (NM\_001979), OCLN  
 CC (NM\_002538), TXNRD1 (NM\_003330), UGT8 (NM\_003360), MRP (NM\_004996),  
 CC NM\_019900, NM\_019901, NM\_019902, NM\_019862, NM\_019898, NM\_019899, and  
 CC their complementary sequences, or a sequence (Si) chosen from 87  
 CC sequences and their complements. The chemical pretreatment is bisulphite  
 CC treatment to convert cytosines (but not methyl-cytosines) into uracils.  
 CC Also included are an oligomer (Ii) in particular an oligonucleotide or a  
 CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one  
 CC base sequence having a length of 9 nucleotides which hybridises to or is  
 CC identical to a chemically pretreated DNA of genes associated with  
 CC pharmacogenomics and their complements, arranged in an array for  
 CC analysing diseases associated with the methylation state (Cpg) and/or  
 CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The  
 CC oligomers may also be used as PCR primers. The set of 87 nucleic acids  
 CC and their complements is useful for diagnosis and therapy of solid  
 CC tumours and cancer. The present sequence represents one the 87 DNA  
 CC sequences or its complement. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 7025 BP; 2377 A; 20 C; 1169 G; 3458 T; 0 U; 1 Other;  
 Query Match 3.3%; Score 35.4; DB 6; Length 7025;  
 Best Local Similarity 61.3%; Pred. No. 5.6;  
 Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;  
 XX  
 QY 536 TGTCATGCTATATGAATGAATGAAGAGAGGTGTGACGGGTAGTACGAATCTGTATGA 595  
 DB 3990 TTTTAGGTATATTTTAAAGAAAAATAAATATGTTTATAGAAGTATGTATTAATGT 4049  
 QY 596 TAATCAATGTCAGTCATGCTGTAACAGCGGA 628  
 DB 4050 TAATTAACGTTTATTAATGTTTAAAGTGA 4082  
 XX  
 RESULT 15  
 ID AAH16164/c  
 XX AAH16164 standard; cDNA; 2604 BP.  
 AC AAH16164;  
 XX  
 XX 26-JUN-2001 (first entry)  
 XX Human cDNA sequence SEQ ID NO:14937.  
 DE  
 XX

PR 07-APR-2000; 2000DE-01019173.  
 PR 30-JUN-2000; 2000DE-01032529.  
 PR 01-SEP-2000; 2000DE-01043826.  
 XX (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2002-010834/01.  
 XX  
 PT New nucleic acid, useful for diagnosis and therapy of metabolic disease,  
 PT solid tumor and cancers, comprises segment of chemically modified genomic  
 PT sequences of genes associated with metabolism.  
 XX  
 PS Claim 1; Page 119-121; 143pp; English.  
 XX  
 CC The invention relates to a nucleic acid (I) comprising a sequence at  
 CC least 18 bases of a segment of the chemically pretreated DNA of genes  
 CC associated with metabolism such as DUSP2 (NM\_004418), EPHX2 (NM\_001979),  
 CC QDPR (NM\_000320), SGSH (NM\_000199), SHMT2 (NM\_005412), SLC7A2  
 CC (NM\_003046), SLC7A4 (NM\_004173) and TYMS (NM\_001071) (all undefined). (I)  
 CC are useful for diagnosis and therapy of metabolic disease, solid tumours  
 CC and cancers; as primer oligonucleotides for the amplification of DNA  
 CC sequences; for detecting the cytosine methylation state and/or single  
 CC nucleotide polymorphisms (SNPs) in a chemically treated DNA of genes  
 CC associated with metabolism. An array of (i) is useful for ascertaining  
 CC genetic and/or epigenetic parameters for the diagnosis and/or therapy of  
 CC existing diseases or the predisposition to specific diseases by analysing  
 CC cytosine methylations. The method involves chemically treating genomic  
 CC DNA sample by a solution of bisulphite, hydrogen sulphite or disulphite  
 CC such that cytosine bases which are unmethylated at the 5th-position are  
 CC converted to uracil or another base which is dissimilar to cytosine in  
 CC terms of hybridisation behaviour and amplifying fragments of the  
 CC chemically pretreated genomic DNA. The genomic DNA is from cells or  
 CC cellular components which contain DNA, sources of DNA comprising, for  
 CC e.g. cell lines, biopsies, blood, sputum, stool, urine, cerebral-spinal  
 CC fluid, tissue embedded in paraffin such as tissue from eye, intestine,  
 CC kidney, brain, heart, prostate, lung, breast or liver, histologic object  
 CC slides and their combinations. Genetic parameters are mutations in  
 CC particular insertions, deletions, point mutations, inversions and  
 CC polymorphisms of genes associated with metabolism and sequences further  
 CC required for their regulation. Epigenetic parameters are in particular  
 CC cytosine methylations and further chemical modifications of DNA bases of  
 CC genes associated with metabolism. Further epigenetic parameters include  
 CC for e.g. the acetylation of histones which correlates with DNA  
 CC methylation. AAS63306-AAS63373 represent chemically pretreated metabolism  
 CC associated genes, and related primers of the invention  
 XX  
 SQ Sequence 7025 BP; 2377 A; 20 C; 1169 G; 3458 T; 0 U; 1 Other;  
 Query Match 3.3%; Score 35.4; DB 6; Length 7025;  
 Best Local Similarity 61.3%; Pred. No. 5.6;  
 Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;  
 QY 536 TGTCATGCTATATGAATGAATGAAGAGAGGTGTGACGGGTAGTACGAATCTGTATGA 595  
 DB 3990 TTTTAGGTATATTTTAAAGAAAAATAAATATGTTTATAGAAGTATGTATTAATGT 4049  
 QY 596 TAATCAATGTCAGTCATGCTGTAACAGCGGA 628  
 DB 4050 TAATTAACGTTTATTAATGTTTAAAGTGA 4082  
 XX  
 RESULT 14  
 ID AAS63350  
 XX AAS63350 standard; DNA; 7025 BP.  
 AC AAS63350;  
 XX  
 XX 29-JAN-2002 (first entry)  
 XX Chemically pretreated metabolism associated gene #45.  
 DE  
 XX Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;  
 KW solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;  
 KW single nucleotide polymorphism detection; SNP; stool; urine; lung;  
 KW cerebral-spinal fluid; intestine; brain; heart; prostate; breast; DUSP2;  
 KW EPHX2; QDPR; SGSH; SHMT2; SLC7A2; SLC7A4; TYMS; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200176451-A2.  
 PN  
 XX 18-OCT-2001.  
 PD  
 XX 06-APR-2001; 2001WO-EP004016.  
 XX  
 XX 06-APR-2000; 2000DE-01019058.  
 PR



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 08:11:45 ; Search time 2997.29 Seconds  
(without alignments)  
15747.756 Million cell updates/sec

Title: US-10-070-386-2  
Perfect score: 1089  
Sequence: 1 taaactccatctataggg.....actgcaatgagtcgcccg 1089

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_ma.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rnd.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sv.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1089	100.0	1089	6	BD013074	Regulator
C 2	47	4.3	2000	6	AX655393	Sequence
C 3	44.8	4.1	2000	6	AX655393	Sequence
C 4	43	3.9	190760	2	AC109734	Rattus no
C 5	43	3.9	214458	2	AC117338	Rattus no
C 6	42.6	3.9	399	6	AR417479	Sequence
C 7	42.6	3.9	399	6	BD113032	EST and e
C 8	42.6	3.9	274651	2	AC095227	Rattus no
C 9	41.8	3.8	14461	6	AX256438	Sequence
C 10	41.8	3.8	128683	2	AC135793	Oryza sat
C 11	41.2	3.8	10732	6	E32986	Gene encodi
C 12	40.8	3.7	128683	2	AC135793	Oryza sat
C 13	40.4	3.7	220327	2	AC125966	Rattus no
C 14	40.4	3.7	24974	2	AC097706	Rattus no
C 15	40.2	3.7	1292	3	BT0307576	Eyturus t
C 16	40.2	3.7	75106	8	AT11642	Arabidops
C 17	40.2	3.7	113526	9	AL354769	Human DNA
C 18	40.2	3.7	174026	2	AC021864	Homo sapi
C 19	40	3.7	152304	2	BX072562	Danio rer
C 20	39.8	3.7	111477	9	AL451162	Human DNA
C 21	39.8	3.7	143678	9	AC005145	Homo sapi
C 22	39.8	3.7	170368	9	AC117518	Homo sapi
C 23	39.8	3.7	195156	9	AC090660	Homo sapi
C 24	39.8	3.7	198127	2	AF001593	Homo sapi
C 25	39.6	3.6	1141	6	AX083744	Sequence
C 26	39.4	3.6	175374	2	AC129987	Rattus no
C 27	39.4	3.6	254830	2	AC095718	Rattus no
C 28	38.2	3.6	91853	9	AC106810	Homo sapi
C 29	39.2	3.6	112445	9	AL590072	Human DNA
C 30	39	3.6	139189	10	AL935138	Mouse DNA
C 31	39	3.6	223545	2	AC100545	Mus muscu
C 32	39	3.6	318620	2	AC110910	Mus muscu
C 33	38.8	3.6	37797	3	CFT21B4	Z81124 Caenorhabdi
C 34	38.8	3.6	125020	9	AF429315	Homo sapi
C 35	38.6	3.5	16130	2	AC013873	Drosophil
C 36	38.6	3.5	41609	3	AF018451	Caenorhab
C 37	38.6	3.5	164761	3	AC023695	Drosophil
C 38	38.6	3.5	198245	3	AC023701	Drosophil
C 39	38.6	3.5	198966	2	AC120873	Mus muscu
C 40	38.6	3.5	319329	3	AE003432	Drosophil
C 41	38.4	3.5	162097	2	BX284614	Danio rer
C 42	38.4	3.5	167531	5	AL928135	Zebrafish
C 43	38.4	3.5	185624	2	BX322551	Danio rer
C 44	38.4	3.5	214843	5	BX005203	Zebrafish
C 45	38.2	3.5	225020	10	AC109247	Mus muscu

ALIGNMENTS

RESULT 1  
BD013074  
LOCUS  
DEFINITION BD013074 1089 bp DNA linear PAT 02-AUG-2002  
Regulatory sequences and expression system functional in mold  
ACCESSION BD013074  
VERSION BD013074.1 GI:22093263  
KEYWORDS WO 0118219-A/2.  
SOURCE unidentified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 1089)  
Watanabe,M. and Murakami,T.  
AUTHORS Regulatory sequences and expression system functional in mold fungi  
TITLE Patent: WO 0118219-A 2 15-MAR-2001;  
JOURNAL

Pred. No. is the number of results predicted by chance to have a





Db	789	WRCWAGWARMKGRYRMKWKXATRYYYWMMWMTWMMWWRWWSYRWMSGMRGRMSAW	730
Qy	542	GTCTATATGAATGAAGAAGAAGGTGTGACGGGTTAGTAGCAATCTGTATGATATCA	601
Db	729	RYCSRMKCAKTKYASSARWTKRKRYSYRRRWTKRKGWYRYRWSCRMTARMKR	670
Qy	502	ATGGTAGCAGTGATGGTAACA---GGGATCGGATCTAGCACTGCTATGCTCGGTA	657
Db	569	RKWAGSNKSWMTWRGARSWYKYSCKAKKTRMTYSYSTGMYGMYSYKXMSW	610
Qy	658	TGTAATCCTGGCTATGTTCAAAAGGGGACATAGAAAGAAATACCTCAGTGTGACATACG	717
Db	609	TSKSYNGKMTCTWYSKMGSTRSSXGRWSGMSRWYMKRKYRMYMKWKCTWR	550
Qy	718	TAACTCTGTACATCTCACTCCAAATTCGAACTTGGAGAGCATATGAATACTAA	777
Db	549	RCMCTRWGTYTTSRSHMTGTRKARYTKRYMTYKRYKWKWJASTKYWMSWYKCRKSNMR	490
Qy	778	ATGGAATCTCTCATATAAGTGAAGAAACAGACGCCCTTTTATTATGAACAGAAAGGTC	837
Db	489	YGYCKACKCCYACWCAAAYSGMMWYRKYSKWRJASTKYWMSWYKCRKSNMR	430
Qy	838	AAGAAAGCTTTCAAGTCATCAGAGGGGTTCCATCCAGATCATCTTCCCTTGAACCA	897
Db	429	YGCXMTYCSYGMKWTYMGSYKYSRCYKYNMYTKGMMYMYISAYSSMTWYIYA	370
Qy	898	TGTTCTCGCATTCAGAAATCGTAGCGATGGAACCGTCCAGGTTGCGCTCATTCCTCTG	957
Db	369	KYWKYKRRGRTMSWYKSKYKCYTWCYKQVRCYRWKMRKTKYKRCYCWRYAT	310
Qy	958	CGTCCCTTGCATAAATATGATTTACCA-TTTTCTTTGCGAGCGCGGTCAAGTGAGCG	1016
Db	309	CYWCCTYKRGWYSRRSMRTAGKWKRSRWCRSYSWYKMYKWKWKYSYMSYGNARS	250
Qy	1017	ACGTGCCACGCTGGAGTCCACAATGACAGTGGATGCTCATCCACGCC	1065
Db	249	SGTWSSAAKXTYKGYSTSRRAKWKACRMYSACRYSRTSYCGCSYC	201
RESULT 3			
AX655393			
LOCUS		2000 bp DNA	linear PAT 22-MAR-2003
DEFINITION		Sequence 5263 from Patent WO03000898.	
ACCESSION		AX655393	
VERSION		AX655393.1	GI:29158207
KEYWORDS			
SOURCE		Oryza sativa	
ORGANISM		Oryza sativa	
REFERENCE			
AUTHORS		Chang H.S., Chen W., Cooper B., Glazebrook J., Goff S.A., Hou Y.M., Katagiri P., Quan S., Tao Y., Whitham S., Xie Z., Zhu T. and Zou G.	
TITLE		Plant genes involved in defense against pathogens	
JOURNAL		Patent: WO 03000898-A 5263 03-JAN-2003;	
FEATURES		Syngenta Participations AG (CH)	
source		Location/Qualifiers	
		1. 2000	

```

Source
i.: 2000
/-organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN

Query Match      4.1%; Score 44.8; DB 6; Length 2000;
Best Local Similarity 11.3%; Pred. No. 0.06;
Matches 104; Conservative 398; Mismatches 410; Indels 12; Gaps 4;

Qy 18 CGTTCCTGMAAGCGAGAGCGCCAGAGAGATCACTCTTTGACACGCTTGAGATCT 77
      |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 73 SCRTGSKWSGSGVSCKKCKYKESKWRGGRGEMRSRMWYGRYCARSGRWAGGS 132

```

Qy	78	GATCTGTGTTTCAGTCTCAGTAACCTGTGTGGAAAGTTACATCTCTGGTCTCCTCTCTAC	137
Db	133	GRMGGKSRMYSYMWCTARGCGCKRKESSKGSWGKTCRGARGGSGSSGAKYKSGNS	192
Qy	138	CAGCCCTCCAGGCCAACACCAAGATGTTAGAGTTCCTGCATCTATTATATGGCTCTGGCGA	197
Db	193	KRMWSSCGRGGRSAYSRYGTSRKYGYKWTYYASRRCMAYMTTYSWACSSYT	252
Qy	198	TGATGACATTATGAGGCATGCAAGCATGCTCTACTGCTGCTCTGTGTGGTTAGTTA	257
Db	253	WCRSKRRMNMKWLKMEWSASYGWYSYKMMCTAYKKSYSYKWCYMYRGGGWRGATR	312
Qy	258	CTTAGCTAGACAATATCACAATACAAATGTTGGTTTCCACATGTCAGCTGTTCTACCG	317
Db	313	YWGRTSRMAMWTKMTYGTGYKMGKGGWAGRMMSMCKNWSKACYMRMRMWRMT	372
Qy	318	TAGTCTGAGTGAATGGTAAATGATATATTCAGCTTGACCCCGCAATATTGTAACAGAG	377
Db	373	RRWAKSSRTSRKRRKRWCMRKRYKRWRGYSRMRSSCKRARMKRCRSGRMMKMGRCGM	432
Qy	378	CCAAATGGGTCACTGGCCCCCAGACATGTGGCTATATAAGCTACTGTCTA	432
Db	433	TCRMKSYGMEMRWKSWKRWASIKYKWSRMYRKKCSRTTWGKTTRGMMGTGRCRYKK	492
Qy	433	--GCAATCAGACTTACTCATAGAAGCTCCCCCTATGCTATATAAAATAGTCACACTAC	490
Db	493	RSQMKRCKRRRWRGMYRMWKRYMSARYMRYCARKKYSYSAARKARCYRGGKYWA	552
Qy	491	AACTACGACAGTGTGAATCCGACAGTCTGGTCTCTTGGACATGTGCATGCTCTATATG	550
Db	553	GMWMKRYKMTYKMMWYKRYKSKCSYCRMSYASCMKSAKAGAKMKRKSMSAWSK	612
Qy	551	AATGAAATAGAAGAAGTGTGACGGTGTAGTACGAATCTGTATGATATAATCAATGGTAGCA	610
Db	613	SMRSSRCKRCKLSRSSNXYMMGGMTSGSRMKNKSY-TCYWRKWSGMSKSTCTWYMY	671
Qy	611	GTGATGTAAACAGCGGATCGGATCTAGCACTGCTATGTCCTGGGTATGTAATCTCGGCT	670
Db	672	SKYTYAKYGSYWRVYRAWCMYMRWYYRYRSYNTYMAVYTSSTMAANTGMKYSGRYWT	731
Qy	671	ATGTTTCATAGCGGCGCATAGAAAGATACTCAGTGTCCAGCATACGTAAGCTCTGTA-C	729
Db	732	SWYTKCKGSKYRSNWYTSWWNAKTWQWRATYTRMMWMYRYSKMWYTWCTMMGYWY	791
Qy	730	ATTTCACGTGCAAAATTTCTGAACAATTTGGAGAGCATTAATGAATACTAAATGGAACCTCTC	789
Db	792	WWTYMKRYMWYKCTKTYMYWSATYWTGTWAAWMAKTQWRWGTGAKTRGRARKRYW	851
Qy	790	ATTATAAGTGGAAAAACAGAGCGCCCTTTTATTATGAACAAGACGGCTCAAGACGCTCTTT	849
Db	852	WKNATWCATKRWMTKGRGAK---WAWTWAKAWRKYIYWSWWRAYYYYYYKTRRYTKTCW	908
Qy	850	CAACGTCTACAGAGCGGTTCCTCCAGATCATCTTCCCTTGAACCACTGTTCTCGCAT	909
Db	909	KARWGSWAYHWNWKGSAKWWNWKGGWGHGKTYWYVCTTWKACGBATKVMWCAGWGA	968
Qy	910	CAGAAATCGTAGCGATGGAACCGT	933
Db	969	MYSYSWTRTYMRTWRWNAWSSRT	992

RESULT		4			
ACL109734/c					
LOCUS					
DEFINITION					
	ACL109734	190760 bp	DNA	linear	HTG 08-OCT-2002
	Rattus norvegicus clone CH230-324E22,	*** SEQUENCING IN PROGRESS			
	***, 2 unordered pieces.				
ACCSSION	ACL09734				
VERSION	ACL09734.4	GI_23195415			
KEYWORDS	HTG; HTGS_PHASEI; HTGS_DRAFT;	HTGS_ENRICHED.			
SOURCE	Rattus norvegicus	(Norway rat)			
ORGANISM	Rattus norvegicus				
	Eukaryota; Metazoa;	Chordata;	Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi;	Muridae; Murinae;			

Rattus.  
1 (bases 1 to 190760)  
Muzny, D., Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H., Baidwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Arda, C., Dederich, D., Delgado, O., Danson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, I., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Gracy, M., Guerra, W., Guera, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Hawlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, D., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulsegod, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Mallory, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, B., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pua, L., Pua, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, W., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Steed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steale, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villalba, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.  
Direct Submission  
Unpublished  
2 (bases 1 to 190760)  
Worley, K. C.  
Direct Submission  
Submitted (07-PEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 190760)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (09-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Sep 19, 2002 this sequence version replaced gi:21738234.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Project Information  
Center project name: QGMU  
Center clone name: CH230-324E22  
Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 169268 bases at least Q40  
Consensus quality: 172261 bases at least Q30  
Consensus quality: 173949 bases at least Q20  
Estimated insert size: 209466; sum-of-contigs estimation  
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation  
NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
NOTE: this is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 180102: contig of 180102 bp in length  
1 180103 180202: gap of unknown length  
1 180203 190760: contig of 10558 bp in length.

FEATURES  
source

1..190760  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-324E22"

misc\_feature

1..1725  
/note="wgs end extension  
clone end:Sp6"

misc\_feature

2044..2981  
/note="clone boundary  
clone end:Sp6  
site:MboI"

misc\_feature

end sequence: EXANV35TV"  
179212..180102  
/note="clone boundary  
clone end:T7  
site:MboI"

misc\_feature

end sequence: EXANV35TV"  
180203..181221  
/note="wgs end extension  
clone end:T7"

ORIGIN

Query Match 3.9%; Score 43; DB 2; Length 190760;  
Best Local Similarity 48.9%; Pred No. 0.43;  
Matches 115; Conservative 0; Mismatches 120; Indels 0; Gaps 0;  
QY 652 TGGGTATGTAATCCCTGGCTATGTTCTATAGGGCGCATAGAAAGAAATACCTCAGTGTGAG 711  
Db 17698 TGTGTATTTCATAGCTCTCTTGTGAAATATTGTTGTCAGAAACAGAGAACTAAATATAAG 71639

QY 712 CATACGTAAGCTCTGTACATTTCTGCAATTTCTGAACAACTTGGAGAGCAATTAATGAAA 771  
Db 71638 AGTTCCTTAACTGGGCACTCTCAAGAGATTGTTGTAATGATAGACACAGAGAT 71579  
QY 772 TACTAAATGGAAGCTCCTCATTATAGTGGAAAACAGAGCGCCCTTTTATTATGAAACAGA 831  
Db 71578 TATTAGGATGATGATCAATGATAAGCAAAAACCAAGCTTTCAGTTACAAACCCA 71519

QY 832 AGCGTCAAGACGCTTTTCAACGTCATCAGAGCGGTTCCATCCAGATCATACTTT 886  
Db 71518 TGTTCCTGTAAGTGAAGTCTGAGGGCAATTCATCTCTTTTGCTTCAGACCAACTAT 71464

RESULT 5  
AC117338  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-351P22, \*\*\* SEQUENCING IN PROGRESS  
AC117338 214458 bp DNA linear HTG 11-OCT-2002  
\*\*\*, 2 unordered pieces.  
AC117338.3 GI:23617965  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
ORGANISM  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus  
1 (bases 1 to 214458)  
Muzny, D. Marie; Metzker, M. Lee.; Abramson, S.; Adams, C.; Alder, J.;  
Allen, C.; Allen, H.; Alsbrooks, S.; Amin, A.; Anguiano, D.;  
Anyalebechi, V.; Aoyagi, A.; Ayodeji, M.; Baca, E.; Baden, H.;  
Baldwin, D.; Bandaranaike, D.; Barber, M.; Barnstead, M.; Benahmed, F.;  
Biswal, K.; Blair, J.; Blankenburg, K.; Blyth, P.; Brown, M.;  
Bryant, N.; Buhay, C.; Burch, P.; Burrell, K.; Calderon, E.;  
Cardenas, V.; Carter, K.; Cavazos, I.; Cesar, H.; Center, A.;  
Chacko, J.; Chavez, D.; Chen, G.; Chen, R.; Chen, Y.; Chen, Z.; Chu, J.;  
Cleveland, C.; Cockrell, R.; Cox, C.; Coyle, M.; Cree, A.; D'Souza, L.;  
Devila, M. L.; Davis, C.; Davy-Carroll, L.; De Anda, C.; Dederich, D.;  
Delgado, O.; Denson, S.; Deramo, C.; Ding, Y.; Dinh, H.; Divya, K.;  
Draper, H.; Dugan-Rocha, S.; Dunn, A.; Durbin, K.; Duval, B.; Evans, K.;  
Egan, A.; Escotto, M.; Eugene, C.; Evans, C. A.; Falls, T.; Fan, G.;  
Fernandez, S.; Finley, M.; Flagg, N.; Forbes, L.; Foster, M.; Foster, P.;  
Fraser, C. M.; Gabisi, A.; Ganta, R.; Garcia, A.; Garner, I.; Garza, M.;  
Guebaragoris, E.; Geer, K.; Gill, R.; Grady, M.; Guerra, W.; Guevara, W.;  
Gunaratne, P.; Haaland, W.; Hamill, C.; Hamilton, C.; Hamilton, K.;  
Harvey, Y.; Havlak, P.; Hayes, A.; Henderson, N.; Hernandez, J.;  
Hernandez, R.; Hines, S.; Hladun, S. L.; Hodgson, A.; Hogues, M.;  
Hollins, B.; Howells, S.; Hulyk, S.; Hume, J.; Idlebird, D.; Jackson, A.;  
Jackson, L.; Jacob, L.; Jiang, H.; Johnson, B.; Johnson, R.; Jolivet, A.;  
Karpathy, S.; Kelly, S.; Khan, Z.; King, L.; Kovar, C.;  
Kowis, C.; Kratt, C. L.; Lebow, H.; Levan, J.; Lewis, L.; Li, Z.; Liu, J.;  
Liu, J.; Liu, W.; Liu, X.; London, P.; Longacre, S.; Lopez, J.;  
Lorenshewa, L.; Loulsegod, H.; Lozado, R. J.; Lu, X.; Ma, J.;  
Maheshwari, M.; Mahindartne, M.; Mahmoud, M.; Malloy, K.; Mangum, A.;  
Mangum, B.; Mapa, P.; Martin, K.; Martin, R.; Martinez, E.;  
Mawhinley, S.; McLeod, M. P.; McNeill, T. Z.; Meenen, B.;  
Milosavljevic, A.; Miner, G.; Minja, E.; Montemayor, J.; Moore, S.;  
Morgan, M.; Morris, K.; Morris, S.; Munday, M.; Murphy, M.; Nair, L.;  
Nankervis, C.; Neal, D.; Newton, N.; Nguyen, N.; Norris, S.;  
Nwaokeme, O.; Okwuonu, G.; Olarnpungoon, A.; Pal, S.; Parks, K.;  
Pasternak, S.; Paul, H.; Perez, A.; Perez, L.; Pfannkuch, C.;  
Plopper, P.; Poindexter, A.; Popovic, D.; Primus, E.; Pu, L.-L.;  
Puzo, M.; Quiroz, J.; Rachlin, E.; Reeves, K.; Regier, M. A.; Reigh, R.;  
Reilly, B.; Reilly, M.; Ren, Y.; Reuter, M.; Richards, S.; Riggs, F.;  
Rives, C.; Rodkey, T.; Rojas, A.; Rose, M.; Rose, R.; Ruiz, S. J.;  
Sanders, M.; Savary, G.; Scherer, S.; Scott, G.; Shatsman, S.; Shen, H.;  
Shetty, J.; Shvartsbeyn, A.; Sisson, I.; Sitter, C. D.; Smajs, D.;  
Speed, A.; Sodargen, E.; Song, X.-Z.; Sorelle, R.; Soosa, J.;  
Steinle, M.; Strong, R.; Sutton, A.; Svatek, A.; Tabor, P.; Taylor, C.;  
Taylor, T.; Thomas, N.; Thomas, S.; Tingey, A.; Trejos, Z.; Usmani, K.;  
Valas, R.; Vera, V.; Villasana, D.; Waldron, L.; Walker, B.; Wang, J.;  
Wang, Q.; Wang, S.; Warren, J.; Warren, R.; Wei, X.; White, P.;  
Williams, G.; Willson, R.; Wlezyk, R.; Wooden, H.; Worley, K.;  
Wright, D.; Wright, J.; Wu, J.; Yakub, S.; Yen, J.; Yoon, L.; Yoon, V.;  
Yu, P.; Zhang, J.; Zhou, X.; Zhou, X.; Zhao, S.; Dunn, D.; von  
Niederhausen, A.; Weiss, R.; Smith, D. R.; Holt, R. A.; Smith, H. O.;  
Weinstock, G. and Gibbs, R. A.  
Direct Submission  
Unpublished  
REFERENCE  
2 (bases 1 to 214458)  
Worley, K. C.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Direct Submission  
Submitted (10-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 214458)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Oct 9, 2002 this sequence version replaced gi:21746250.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

Center: Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GUAX  
Center clone name: CH230-351P22  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 145461 bases at least Q40  
Consensus quality: 14349 bases at least Q30  
Consensus quality: 147003 bases at least Q20  
Estimated insert size: 156806; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: this is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 212885: contig of 212885 bp in length  
212886 212985: gap of unknown length  
212986 214458: contig of 1473 bp in length.

Location/Qualifiers  
1. 214458  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-351P22"  
1. 1014  
/note="wgs end extension  
clone end:Sp6"  
39152..40502  
/note="wgs end extension  
clone end:Sp6"  
44260..45151  
/note="clone boundary  
clone end:Sp6  
site:Mbol  
end sequence:RXASL95TV"  
complement(210250..210949)  
/note="clone boundary  
clone end:T7  
site:Mbol

source  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature

end\_sequence:RXASL95JT"
ORIGIN
Query Match 3.9%; Score 43; DB 2; Length 214458;
Best Local Similarity 48.9%; Pred. No. 0.43;
Matches 115; Conservative 0; Mismatches 120; Indels 0; Gaps 0;
QY 652 TGGGTATGTAATCCTGGCTATGTTTCATAGGGCGACATAGAGAAATACCTCAGTGTGAG 711
Db 63587 TGGGTATGTAATCCTGGCTATGTTTCATAGGGCGACATAGAGAAATACCTCAGTGTGAG 711
QY 712 CATACGTAAGCTCTGTACATCTTCACCTGCAATTTCTGAACAATTTGGAGAGCAATATGAAA 771
Db 63647 AGTTCCTTAATCCTGGCCATCTCATGAGATTGATTGATAATGATAGACACAGAAT 63706
QY 772 TACTAATGGAATCTCTCATATAGTGGAAACAGAGCGCCCTTTTNTATGAAACAGA 831
Db 63707 TATTAGCGATGAGTCAACATGATAAAGCAAAACACCAAGCTTTTCAGTTACAAACCCA 63766
QY 832 AGCGTCAAGAACCTCTTTCAACGCTCATCAGAGCGGTTCCATCCAGATCATACTTTT 886
Db 63767 TGTTTCTGTAGTGAATCTCAGCGGCAATCATCTTTTGCTCAGACCAACTAT 63821
RESULT 6
AR417479
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
ORIGIN
Query Match 3.9%; Score 42.6; DB 6; Length 399;
Best Local Similarity 9.6%; Pred. No. 0.22;
Matches 33; Conservative 164; Mismatches 148; Indels 0; Gaps 0;
QY 740 AATTTCTGAACAATTTGAGAGCATTATGAATATACTAAATGAATGAACTCTCTTATAAGTG 799
Db 16 RAMTRAYWMTYRGSKSWRAKMSWKRRRRRAMWKKSWKSWKSWKSWKSWKSWKSWKSWK 75
QY 800 GAAACACAGAGCGCCCTTTTATTATGAACAGAGCGTCAAGAACGTCCTTTCAACGTCATC 859
Db 76 RMKGRGAASWAGYMSMTYMTFRWRYRYRKKACTKWRRAAGWAGWAGWAWAYAKWYMA 135
QY 860 AGAGCGGTTCCATCCAGATCATCTTCCCTTGACCATGTTCTCGCATTCAGATCGTA 919
Db 136 WRTAMKYNNMKSKRSRRRRRWYNNYNNRRRTWNNRRASCYRGAAYASAGYNNWNN 195
QY 920 GCGATGAAACCGTCCAGGGTTCCTGTCATTCCTTCCTGCTCCCTTGCAATAAATCGTA 979
Db 196 YNNRRKNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNN 255
QY 980 TTACCATTTTCTTCGAGCGCGGTCACGTCAGCAGCGTCCACGTCGAGTGCACCA 1039
Db 256 YRRSCCWNNSCFWTKYRNSWTSCASYISYKTRASCMMCCMMKWRMMWMMWMMWMMWMM 315
QY 1040 ATGACGAGTGGATCGTCAATCCAGCCACTCACTGCAATGAGTCGC 1084
Db 316 CKKTSAMRYRSCYSAKNNRRWMMCCAYNNKTCNNWMMWMMWMMWMMWMMWMMWMMWMM 360
RESULT 7
BD113032
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
ORIGIN
Query Match 3.9%; Score 42.6; DB 6; Length 399;
Best Local Similarity 9.6%; Pred. No. 0.22;
Matches 33; Conservative 164; Mismatches 148; Indels 0; Gaps 0;
QY 740 AATTTCTGAACAATTTGAGAGCATTATGAATATACTAAATGAATGAACTCTCTTATAAGTG 799
Db 16 RAMTRAYWMTYRGSKSWRAKMSWKRRRRRAMWKKSWKSWKSWKSWKSWKSWKSWKSWK 75
QY 800 GAAACACAGAGCGCCCTTTTATTATGAACAGAGCGTCAAGAACGTCCTTTCAACGTCATC 859
Db 76 RMKGRGAASWAGYMSMTYMTFRWRYRYRKKACTKWRRAAGWAGWAGWAWAYAKWYMA 135
QY 860 AGAGCGGTTCCATCCAGATCATCTTCCCTTGACCATGTTCTCGCATTCAGATCGTA 919
Db 136 WRTAMKYNNMKSKRSRRRRRWYNNYNNRRRTWNNRRASCYRGAAYASAGYNNWNN 195
QY 920 GCGATGAAACCGTCCAGGGTTCCTGTCATTCCTTCCTGCTCCCTTGCAATAAATCGTA 979
Db 196 YNNRRKNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNN 255
QY 980 TTACCATTTTCTTCGAGCGCGGTCACGTCAGCAGCGTCCACGTCGAGTGCACCA 1039
Db 256 YRRSCCWNNSCFWTKYRNSWTSCASYISYKTRASCMMCCMMKWRMMWMMWMMWMMWMM 315
QY 1040 ATGACGAGTGGATCGTCAATCCAGCCACTCACTGCAATGAGTCGC 1084
Db 316 CKKTSAMRYRSCYSAKNNRRWMMCCAYNNKTCNNWMMWMMWMMWMMWMMWMMWMMWMM 360
RESULT 8
AC095227
LOCUS
DEFINITION
ACCESSION

end\_sequence:RXASL95JT"
ORIGIN
Query Match 3.9%; Score 43; DB 2; Length 214458;
Best Local Similarity 48.9%; Pred. No. 0.43;
Matches 115; Conservative 0; Mismatches 120; Indels 0; Gaps 0;
QY 652 TGGGTATGTAATCCTGGCTATGTTTCATAGGGCGACATAGAGAAATACCTCAGTGTGAG 711
Db 63587 TGGGTATGTAATCCTGGCTATGTTTCATAGGGCGACATAGAGAAATACCTCAGTGTGAG 711
QY 712 CATACGTAAGCTCTGTACATCTTCACCTGCAATTTCTGAACAATTTGGAGAGCAATATGAAA 771
Db 63647 AGTTCCTTAATCCTGGCCATCTCATGAGATTGATTGATAATGATAGACACAGAAT 63706
QY 772 TACTAATGGAATCTCTCATATAGTGGAAACAGAGCGCCCTTTTNTATGAAACAGA 831
Db 63707 TATTAGCGATGAGTCAACATGATAAAGCAAAACACCAAGCTTTTCAGTTACAAACCCA 63766
QY 832 AGCGTCAAGAACCTCTTTCAACGCTCATCAGAGCGGTTCCATCCAGATCATACTTTT 886
Db 63767 TGTTTCTGTAGTGAATCTCAGCGGCAATCATCTTTTGCTCAGACCAACTAT 63821
RESULT 6
AR417479
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
ORIGIN
Query Match 3.9%; Score 42.6; DB 6; Length 399;
Best Local Similarity 9.6%; Pred. No. 0.22;
Matches 33; Conservative 164; Mismatches 148; Indels 0; Gaps 0;
QY 740 AATTTCTGAACAATTTGAGAGCATTATGAATATACTAAATGAATGAACTCTCTTATAAGTG 799
Db 16 RAMTRAYWMTYRGSKSWRAKMSWKRRRRRAMWKKSWKSWKSWKSWKSWKSWKSWKSWK 75
QY 800 GAAACACAGAGCGCCCTTTTATTATGAACAGAGCGTCAAGAACGTCCTTTCAACGTCATC 859
Db 76 RMKGRGAASWAGYMSMTYMTFRWRYRYRKKACTKWRRAAGWAGWAGWAWAYAKWYMA 135
QY 860 AGAGCGGTTCCATCCAGATCATCTTCCCTTGACCATGTTCTCGCATTCAGATCGTA 919
Db 136 WRTAMKYNNMKSKRSRRRRRWYNNYNNRRRTWNNRRASCYRGAAYASAGYNNWNN 195
QY 920 GCGATGAAACCGTCCAGGGTTCCTGTCATTCCTTCCTGCTCCCTTGCAATAAATCGTA 979
Db 196 YNNRRKNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNN 255
QY 980 TTACCATTTTCTTCGAGCGCGGTCACGTCAGCAGCGTCCACGTCGAGTGCACCA 1039
Db 256 YRRSCCWNNSCFWTKYRNSWTSCASYISYKTRASCMMCCMMKWRMMWMMWMMWMMWMM 315
QY 1040 ATGACGAGTGGATCGTCAATCCAGCCACTCACTGCAATGAGTCGC 1084
Db 316 CKKTSAMRYRSCYSAKNNRRWMMCCAYNNKTCNNWMMWMMWMMWMMWMMWMMWMMWMM 360
RESULT 7
BD113032
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
ORIGIN
Query Match 3.9%; Score 42.6; DB 6; Length 399;
Best Local Similarity 9.6%; Pred. No. 0.22;
Matches 33; Conservative 164; Mismatches 148; Indels 0; Gaps 0;
QY 740 AATTTCTGAACAATTTGAGAGCATTATGAATATACTAAATGAATGAACTCTCTTATAAGTG 799
Db 16 RAMTRAYWMTYRGSKSWRAKMSWKRRRRRAMWKKSWKSWKSWKSWKSWKSWKSWKSWK 75
QY 800 GAAACACAGAGCGCCCTTTTATTATGAACAGAGCGTCAAGAACGTCCTTTCAACGTCATC 859
Db 76 RMKGRGAASWAGYMSMTYMTFRWRYRYRKKACTKWRRAAGWAGWAGWAWAYAKWYMA 135
QY 860 AGAGCGGTTCCATCCAGATCATCTTCCCTTGACCATGTTCTCGCATTCAGATCGTA 919
Db 136 WRTAMKYNNMKSKRSRRRRRWYNNYNNRRRTWNNRRASCYRGAAYASAGYNNWNN 195
QY 920 GCGATGAAACCGTCCAGGGTTCCTGTCATTCCTTCCTGCTCCCTTGCAATAAATCGTA 979
Db 196 YNNRRKNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNN 255
QY 980 TTACCATTTTCTTCGAGCGCGGTCACGTCAGCAGCGTCCACGTCGAGTGCACCA 1039
Db 256 YRRSCCWNNSCFWTKYRNSWTSCASYISYKTRASCMMCCMMKWRMMWMMWMMWMMWMM 315
QY 1040 ATGACGAGTGGATCGTCAATCCAGCCACTCACTGCAATGAGTCGC 1084
Db 316 CKKTSAMRYRSCYSAKNNRRWMMCCAYNNKTCNNWMMWMMWMMWMMWMMWMMWMMWMM 360
RESULT 8
AC095227
LOCUS
DEFINITION
ACCESSION













(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GRGV  
Center clone name: CH230-2303  
----- Summary Statistics  
Assembly program: Atlas 3.0  
Consensus quality: 203711 bases at least Q40  
Consensus quality: 206214 bases at least Q30  
Consensus quality: 207905 bases at least Q20  
Estimated insert size: 211787; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 216855: contig of 216855 bp in length  
\* 216856: gap of unknown length  
\* 216956 218421: contig of 1466 bp in length  
\* 218422 218521: gap of unknown length  
\* 218522 220327: contig of 1806 bp in length.  
----- Location/Qualifiers  
FEATURES  
source  
1..220327  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-2303"  
misc\_feature  
1..1645  
/notes="wgs end extension  
clone\_end:T7"  
misc\_feature  
4726..6822  
/notes="wgs end extension  
clone\_end:T7"  
misc\_feature  
complement(7464..8057)  
/notes="clone boundary  
clone\_end:T7"  
misc\_feature  
site:ECORI  
end\_sequence:BH275795"  
misc\_feature  
complement(215393..215875)  
/notes="clone boundary  
clone\_end:Sp6  
site:ECORI  
end\_sequence:BH275796"  
-----  
ORIGIN  
Query Match 3.7%; Score 40.4; DB 2; Length 220327;  
Best Local Similarity 53.1%; Pred. No. 2.7;  
Matches 86; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
-----  
QY 211 TGAGGATGACGACATGCTACTGCTCTGTTGGTTAGGTACCTTACCTAGCTAGACA 270  
Db 14182 TCAAGCTCTGATACATGCTAGTGGTGTATACCTTCATATCCATCCATACC 14123

QY 271 ATATCACAATACAAATGTTGTTCCACATGTCAGCTGGTCTTACCTAGCTAGTGA 330  
Db 14122 TAAGCAATAGTAATGTTGTTGACATGGCATGTTGATTAACCAACATGAGTCTA 14063  
QY 331 ATGGGTAATGATATATGAGCTTGACCCCGCAATATGTAA 372  
Db 14062 TTGTCTCTGATGTTAGGCGAGCAATCTGCCCTATGATA 14021  
RESULT 14  
AC097706 244974 bp DNA linear HTG 13-MAY-2003  
LOCUS Rattus norvegicus clone CH230-72A1, WORKING DRAFT SEQUENCE, 4  
DEFINITION unordered pieces.  
AC097706  
AC097706.7 GI:30580908  
HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE  
1 (bases 1 to 244974)  
Mizny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandazanaika, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyie, M., Cree, A., D'Souza, L., Davala, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Ducan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Louisedge, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E. E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaekelemeh, O., Okwuonu, G., Olarnpungsoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Speed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, N., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleciyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.  
DIRECT SUBMISSION

TITLE

Matches	86;	Conservative	0;	Mismatches	76;	Indels	0;	Gaps	0;
QY	211	TAAGCATGCACGACGAGCTCTACTGCTGCTCTGTTGGTTAGGTTACCTTAGCTAGACA	270						
Ddb	25204	TCAAAGCCTCTCATATCAATAAAGCTGTAGTGTAGCTGTTTAAACCCCTCATATCCATGCCATACC	25263						
QY	271	ATATCACAAATACAAAATGTGGTTTTCCACATGTCAGCTGGTTCTACCGTAGTCTCAGATGAA	330						
Ddb	25264	TAAACACAAATTAGTAAATGTTTGTGCATATGGCATGTTGATTAAACCAACTGAGTCTA	25323						
QY	331	ATGGGTAAATTCATATATTTAGCTTTGACCCCGCAATATTTGTA	372						
Ddb	25324	TTTGTCTCTGTATGTGTAGGGCAGCAATCTGCCCTATGATAA	25365						
RESULT 15									
LOCUS	BT0307576	1292 bp	DNA	linear	INV 10-MAR-2001				
DEFINITION	Bycurus tomentosus mitochondrial COI and COII genes, strain Tom16.								
ACCESSION	AJ307576								
VERSION	AJ307576.1	GI:13276113							
KEYWORDS	COI gene; COII gene.								
ORGANISM	mitochondrion Bycurus tomentosus								
	Bycurus tomentosus								
	Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;								
	Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;								
	Bycuridae; Byturus.								
	1 (bases 1 to 1292)								
REFERENCE									
AUTHORS	Malloch,G.L.								
TITLE	Direct Submmission								
JOURNAL	Submitted (01-MAR-2001) Malloch G.L., Vegetative systems, Scottish								
	Crop Research Institute, SCRI, Invergowrie, Dundee, Tayside.								

```

REFERENCE 1 (bases 1 to 1292)
AUTHORS Malloch, G.L.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2001) Malloch G.L., Vegetative systems, Scottish
Crop Research Institute, SCRI, Invergowrie, Dundee, Tayside,
Scotland, DD2 5DA, UK

FEATURES             Location/Qualifiers
     source            1..1292
                        /organism="Byturus tomentosus"
                        /organelle="mitochondrion"
                        /mol_type="genomic DNA"
                        /strain="Tom16"
                        /isolate="16"
                        /db_xref="taxon:153020"
     gene              1..771
                        /gene="COI"
     misc_feature      <1..>771
                        /gene="COI"
     misc_feature      772..833
                        /note="intermediate"
     gene              834..1292
                        /gene="COII"
     misc_feature      834..>1292
                        /gene="COII"
ORIGIN

```

Query Match	3.7%	Score 40.2	DB 3	Length 1292
Best Local Similarity	50.3%	Pred. No. 1.5	98	Indels 0
Matches 99	Conservative 0	Mismatches		Gaps 0
802	AAACAGAGCGCCCTTTATTATGAAACAGAGAGCGTCAAGAACGCTCTTTTCAACGTCATCAG	861		
Ddb	808	ATAAAGATGGAACCTTTTATTAGGAATGGCAACGTGAAGATCTACTTTTACTTCAAGATAG	867	
Qy	862	AGCGGTTCCATCCAGATCATCTTTCCCTTGAACCATGTTCTCGCATTCAGATCGTAGC	921	
Ddb	868	AAGATCTCCATTAATAGAACACTTTCTTTTTCATGATCATGCTTTACTAGTATTAGT	927	
Qy	922	GATGAAACCGTCAGGGTTGCGTGTCATTCCTTCGTTGCGTCCCTTGCATATAAAATCGTATT	981	
Ddb	928	AATTATTACTATTTTAGTAGGCCAATTAATAGTTAGCTTATTTTATAATAATACAAA	987	
Qy	982	ACCATTTTCTTTCCGAG	998	
Ddb	988	TGCTATTATTAGAAG	1004	

Search completed: June 10, 2004, 11:49:43  
Job time : 3003.79 secs

